

## SEQUENCE LISTING

## (1) GENERAL INFORMATION:

- (i) APPLICANT: HUSE, WILLIAM D.
- (ii) TITLE OF INVENTION: SURFACE EXPRESSION LIBRARIES OF HETEROMERIC RECEPTORS
- (iii) NUMBER OF SEQUENCES: 75
- (iv) CORRESPONDENCE ADDRESS:
  - (A) ADDRESSEE: PRETTY, SCHROEDER, BRUEGEMANN & CLARK
  - (B) STREET: 444 SO. FLOWER STREET, SUITE 200
  - (C) CITY: LOS ANGELES
  - (D) STATE: CALIFORNIA
  - (E) COUNTRY: UNITED STATES
  - (F) ZIP: 90071
- (v) COMPUTER READABLE FORM:
  - (A) MEDIUM TYPE: Floppy disk
  - (B) COMPUTER: IBM PC compatible
  - (C) OPERATING SYSTEM: PC-DOS/MS-DOS
  - (D) SOFTWARE: PatentIn Release #1.0, Version #1.25
- (vi) CURRENT APPLICATION DATA:
  - (A) APPLICATION NUMBER:
  - (B) FILING DATE:
  - (C) CLASSIFICATION:
- (viii) ATTORNEY/AGENT INFORMATION:
  - (A) NAME: CAMPBELL, CATHRYN A.
  - (B) REGISTRATION NUMBER: 31,815
  - (C) REFERENCE/DOCKET NUMBER: P31 8882
- (ix) TELECOMMUNICATION INFORMATION:
  - (A) TELEPHONE: 619-535-9001
  - (B) TELEFAX: 619-535-8949

## (2) INFORMATION FOR SEQ ID NO:1:

- (i) SEQUENCE CHARACTERISTICS:
  - (A) LENGTH: 7445 base pairs
  - (B) TYPE: nucleic acid
  - (C) STRANDEDNESS: both
  - (D) TOPOLOGY: circular

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO:1:

AATGCTACTA CTATTAGTAG AATTGATGCC ACCTTTTCAG CTCGGCCCC AAATGAAAAT	60
ATAGCTAAAC AGTTTATTGA CCATTTGCGA AATGTATCTA ATGGTCAAAC TAAATCTACT	120
CGTTCCGAGA ATTGGGAATC AACTGTTACA TCCAATGAAA CTTCCAGACA CCGTACTTTA	180
GTTGCATATT TAAAACATGT TGAGCTACAG CACCAGATTG AGCAATTAAG CTCTAAGCCA	240
TCTGCAAAAA TGACCTCTTA TCAAAAGGAG CAATTAAAGG TACTCTCTAA TCCTGACCTG	300
TTGGAGTTTG CTTCCGGTCT GGTTCGCTT GAAGCTCGAA TTAAAACCGG ATATTTGAAG	360
TCTTTGGGC TTCCTCTTAA TCTTTTGAT GCAATCCGCT TTGCTTCTGA CTATAATAGT	420

CAGGGTAAAG ACCTGATTIT TCATTTATGG TCATTCTCGT TTTCTGAAC	480
TTTGAGGGGG ATTCAATGAA TATTTATGAC GATTCCGCAG TATTGCCAC	540
AAACATTITA CTATTACCCC CTCTGGAAA ACTTCCTTTG CAAAAGCCTC	600
GGTTTTATC GTCGTCTGGT AAACGAGGGT TATGATAGTG TTGCTCTTAC	660
AATTCCCTTT GGCGTTATGT ATCTGCATTA GTTGAATGTG GTATTCTAA	720
ATCTCAACTG ATGAATCTTT CTACCTGTAA TAATGTGTT CCGTTAGTTC	780
TCTTCCCAAC GTCCTGACTG GTATAATGAG CCAGTTCTTA AAATGCCATA	840
CAATGATTAA AGTTGAAATT AAACCATCTC AAGCCCATT TACTACTCGT	900
CTCGTCAGGG CAAGCCTTAT TCACTGAATG AGCAGCTTG TTACGTTGAT	960
TTGGGTAATG AATATCCGGT TCTTGTCAAG ATTACTCTTG ATGAAGGTCA	1020
GCCAGCCTAT GCGCCTGGTC TGTACACCGT TCATCTGTCC TCTTCAAAG	1080
TTGGTCAGTT CGGTTCCCTT ATGATTGACC GTCTGCGCCT CGTICCGGCT	1140
AAAGTAACATG GAGCAGGTCC CGGATTTCGA CACAATTAT CAGGCCATGA	1200
TACAAATCTC CGTTGTACTT TGTTTGCAGC TTGGTATAAT CGCTGGGGGT	1260
CAAAGATGAG TGTTTAGTG TATTCTTTCG CCTCTTCTGT TTTAGGGTGG	1320
TGCCTTCGTA GTGGCATTAC GTATTTTACCGT TTAACTCCCT GCAAGCCTCA	1380
GCGACCCAAT ATATCGTTA TGCGTGGGCG ATGGTTGTTG TCATTGTCGG	1440
CGCAACTATC GGTATCAAGC TGTTTAAGAA ATTCACTCTG AAAGCAAGCT	1500
GATAAACCGA TACAATTAAA GGCTCCCTTT GGAGCCTTT TTTTGGAGA	1560
TTTCAACGT GAAAAAATTA TTATTCGAA TTCCTTGT TGTTCCCTTC	1620
TATTCTCACT CCGCTGAAAC TGTTGAAAGT TGTTAGCAA AACCCCATAC	1680
AGAAAATICA TTACTAACG TCTGGAAAGA CGACAAAATC TTAGATCGTT	1740
ACGCTAACTA TGAGGGTTGT CTGTGGAATG CTACAGGCGT TGTAGTTGT	1800
ACTGGTGACG AAACTCAGTC TTACGGTACA TGGGTTCCCA TTGGGCTTGC	1860
TATCCCTGAA ATGAGGGTG GTGGCTCTGA GGGTGGCGGT ACTAAACCTC	1920
CTGAGTACGG TGATACACCT ATTCGGGCT ATACTTATAT CAACCCCTTC	1980
GACGGCACTT ATCCCTAATCC TTCTCTTGAG GAGTCTCAGC CTCTTAAATAC	2040
TTTCATGTTT CAGAATAATA GGTTCCGAAA TAGGCAGGGG GCATTAAC	2100
ACTG TTTATACGGG CACTGTTACT CAAGGCACG ACCCCGTTAA	2160
AACTTATTAC CAGTACACTC CTGTATCATC AAAAGCCATG TATGACGCTT	2220
ACTGGAACGG TAAATTCAGA GACTGCGCTT TCCATTCTGG CTTTAATGAA	2280
GATCCATTCC TTTGTGAATA TCAAGGCCAA TCGTCTGACC TGCCTCAACC	2340
TCCCTGTCAAT GCTGGCGCG GCTCTGGTGG TGTTCTGGT GGCGGCTCTG	2400
AGGGTGGTGG CTCTGAGGGGA GGCGGTTCCG GTGGTGGCTC TGTTCCGGT	2460
GATTTGATT ATGAAAAGAT GGCAAACGCT AATAAGGGGG CTATGACCGA AAATGCCGAT	

GAAAACGGCC TACAGTCTGA CGCTAAAGGC AAACTTGATT CTGTCGCTAC TGATTACGGT 2520  
 GCTGCTATCG ATGGTTTCA TCGTGACGTT TCCGGCCTTG CTAATGGTAA TGGTGCTACT 2580  
 GGTGATTTTG CTGGCTCTAA TTCCCAAATG GCTCAAGTCG GTGACGGTGA TAATTACCT 2640  
 TTAATGAATA ATTTCCGTCA ATATTTACCT TCCCTCCCTC AATCGGTGA ATGTCGCCCT 2700  
 TTGTCCTTA GCGCTGGTAA ACCATATGAA TTTTCTATTG ATTGTGACAA AATAAACCTA 2760  
 TTCCGTGGTG TCTTTGCGTT TCTTTTATAT GTGCCACCT TTATGTATGT ATTTTCTACG 2820  
 TTGCTAACCA TACTGCGTAA TAAGGAGTCT TAATCATGCC AGTTCTTTTG GGTATCCGT 2880  
 TATTATTCGG TTTCCCTCGGT TTCCCTCTGG TAACTTGTT CGGCTATCTG CTTACTTTTC 2940  
 TTAAAAAGGG CTTCGGTAAG ATAGCTATTG CTATTCATT GTTCTTGCT CTTATTATTG 3000  
 GGCTTAACTC AATTCTTGTG GGTTATCTCT CTGATATTAG CGCTCAATTAA CCCTCTGACT 3060  
 TTGTCAGGG TGTTCAGTTA ATTCTCCGT CTAATGGCT TCCCTGTTT TATGTTATTG 3120  
 TCTCTGTAAA GGCTGCTATT TTCACTTTTG ACGTTAAACA AAAAATCGTT TCTTATTG 3180  
 ATTGGGATAAA ATAATATGGC TGTAACTGGCA AATTAGGCTC TGGAAAGACG 3240  
 CTCGTTAGCG TTGGTAAGAT TCAGGATAAA ATTGTAGCTG GGTGCAAAAT AGCAACTAAT 3300  
 CTGATTAA GGCTCAAAA CCTCCCCAA GTCGGGAGGT TCGCTAAAC GCCTCGCGTT 3360  
 CTTAGAATAC CGGATAAGCC TTCTATATCT GATTGCTTG CTATTGGCG CGGTAATGAT 3420  
 TCCTACGATG AAAATAAAAA CGGCTTGCTT GTTCTCGATG AGTGGGCTAC TTGGTTAAAT 3480  
 ACCCGTTCTT GGAATGATAA GGAAAGACAG CCGATTATTG ATTGGTTCT ACATGCTCGT 3540  
 AAATTAGGAT GGGATATTAT TTTTCTGTT CAGGACTTAT CTATTGTTGA TAAACAGGGCG 3600  
 CGTTCTGCAT TAGCTGAACA TGTGTTTAT TGTGTCGTC TGGACAGAAAT TACTTTACCT 3660  
 TTGTCGGTA CTTTATATTC TCTTATTACT GGCTCGAAAA TGCCTCTGCC TAAATTACAT 3720  
 GTTGGCGTTG TTAAATATGG CGATTCTCAA TTAAGCCCTA CTGTTGAGCG TTGGCTTAT 3780  
 ACTGGTAAGA ATTGTATAA CGCATATGAT ACTAACACAGG CTTTTCTAG TAAATTATGAT 3840  
 TCCGGTGTIT ATTCTTATTT AACGCCCTAT TTATCACACGG GTCGGTATTT CAAACCATTAA 3900  
 AATTAGGTC AGAAGATGAA GCTTACTAAA ATATATTGA AAAAGTTTTC ACGCGTTCTT 3960  
 TGTCTTGGCA TTGGATTGCA ATCAGCATT ACATATAGTT ATATAACCCA ACCTAAGCCG 4020  
 GAGGTTAAAA AGTAGTCTC TCAGACCTAT GATTTGATA AATTCACTAT TGACTCTCT 4080  
 CAGCGTCTTA ATCTAAGCTA TCGCTATGTT TTCAAGGATT CTAAGGGAAA ATTAATTAAT 4140  
 AGCGACCAATT TACAGAAGCA AGGTATTCA CTCACATATA TTGATTATG TACTGTTTCC 4200  
 ATTAAAAAG GTAATTCAAA TGAAATTGTT AAATGTAATT AATTGTTTT TCTTGATGTT 4260  
 TGTTCATCA TCTTCTTTG CTCAGGTAAT TGAAATGAAT AATTGCCCTC TCCGGCGATT 4320  
 TGTAACCTGG TATTCAAAGC AATCAGGGCA ATCCGTTATT GTTCTCCCG ATGTAAGG 4380  
 TACTGTTACT GTATATTCA CTGACGTTAA ACCTGAAAAT CTACGCAATT TCTTATTC 4440  
 TGTGTTACGT GCTAATAATT TTGATATGGT TGGTCAATT CCTTCCATAA TTCAGAAGTA 4500

TAATCCAAAC AATCAGGATT ATATTGATGA ATTGCCATCA TCTGATAATC AGGAATATGA	4560
TGATAATTCC CCTCCTTCTG GTGGTTTCCTT TGTTCCGCAA AATGATAATG TTACTCAAAC	4620
TTTTAAAATT AATAACGTTG GGGCAAAGGA TTTAATACGA GTTGTGAAAT TGTTTGAA	4680
GTCTAATACT TCTAAATCCT CAAATGTATT ATCTATTGAC GGCTCTAATC TATTAGTTGT	4740
TAGTGCACCT AAAGATATT TAGATAACCT TCCTCAATTG CTTTCTACTG TTGATTGCC	4800
AACTGACCAG ATATTGATTG AGGGTTTGAT ATTTGAGGTG CAGCAAGGTG ATGCTTAA	4860
TTTTCAATTG GCTGCTGGCT CTCAGCGTGG CACTGTTGCA GGCGGTGTTA ATACTGACCG	4920
CCTCACCTCT GTTTTATCTT CTGCTGGTGG TTGTTGCGT ATTTTAATG GCGATGTTT	4980
AGGGCTATCA GTTGCAGCAT TAAAGACTAA TAGCCATTCA AAAATATTGT CTGTGCCACG	5040
TATTCTTACG CTTTCAGGTC AGAAGGGTTC TATCTCTGTT GGCCAGAATG TCCCTTTAT	5100
TACTGGTCTG GTGACTGGTG AATCTGCAA TGAAATAAT CCATTTCAGA CGATTGAGCG	5160
TCAAAATGTA GGTATTCCA TGAGCGTTT TCCTGTTGCA ATGGCTGGCG GTAATATTGT	5220
TCTGGATAATT ACCAGCAAGG CCGATAGTTT GAGTTCTCT ACTCAGGCAA GTGATGTTAT	5280
TACTAACAA AGAAGTATTG CTACAACGGT TAATTGCGT GATGGACAGA CTCTTTACT	5340
CGGTGGCCTC ACTGATTATA AAAACACTTC TCAAGATTCT GGCGTACCGT TCCTGTCATA	5400
AATCCCTTA ATCGGCCTCC TGTTTAGCTC CCGCTCTGAT TCCAACGAGG AAAGCACGTT	5460
ATACGTGCTC GTCAAAGCAA CCATAGTACG CGCCCTGTAG CGGGCATTAA AGCGCGGGGG	5520
GTGTGGTGGT TACGGCGAGC GTGACCGCTA CACTGCGAG CGCCCTAGCG CCCGCTCCTT	5580
TCGGTTCTT CCCTTCCTT CTCGCCACGT TCGCCCCCTT TCCCCGTCAA CCTCTAAATC	5640
GGGGGCTCCC TTTAGGGTTC CGATTTAGTG CTITACGGCA CCTCGACCCC AAAAAACTTG	5700
ATTTGGGTGA TGGTTCACGT AGTGGGCCAT CGCCCTGATA CACGGTTTTT CGCCCTTTGA	5760
CGTTGGAGTC CACGGTCTT AATAGTGGAC TCTGTTCCA AACTGGAACA ACACCTCAACC	5820
CTATCTCGGCTTATTCTTT GATTATAAG GGATTTGCC GATTICGAA CCACCATCAA	5880
ACAGGATTTC CGCCTGCTGG GGCAAACCG CGTGGACCGC TTGCTGAAAC TCTCTCAGGG	5940
CCAGGCCGTG AAGGGCAATC AGCTGTTGCC CGTCTCGCTG GTGAAAAGAA AAACCCACCT	6000
GGCGCCCAAT ACGCAAACCG CCTCTCCCCG CGCGTTGGCC GATTCAATTAA TCGAGCTGGC	6060
ACGACAGGTG TCCCAGCTGG AAAGCGGGCA GTGAGCGCAA CGCAATTAAAT GTGAGTTAGC	6120
TCACTCAATTAA GGCAACCCAG GCTTACACT TTATGTTCC GGCTCGTATG TTGTGTGGAA	6180
TTGTGAGCGG ATAACAATTT CACACGGCTC ACTTGGCACT GGCGTCTGTT TTACAACGTC	6240
GTGACTGGGA AAACCTGGC GTTACCCAAG CTTTGTACAT GGAGAAAATA AAGTGAACAA	6300
AAGCACTATT GCACTGGCAC TCTTACCGTT ACCGTTACTG TTTACCCCTG TGACAAAAGC	6360
CGCCCAGGTG CAGCTGCTCG AGTCAGGCCT ATTGTGCCA GGGGATTGTA CTAGTGGATC	6420
CTAGGCTGAA GGCGATGACC CTGCTAAGGC TGCATTCAAT AGTTTACAGG CAAGTGCTAC	6480
TGAGTACATT GGCTACGCTT GGGCTATGGT AGTAGTTATA GTTGGTGCTA CCATAGGGAT	6540

TAAATTATTC AAAAGTTTA CGAGCAAGGC TTCTTAAGCA ATAGCCAAGA GGCCCCCACC 6600  
 GATCCGCCCTT CCCAACAGTT GCGCAGCCTG AATGGCAAT GGGGCTTTCG CTGGTTTCCG 6660  
 GCACCAAG CGGTGCCGG AAGCTGGCTG GAGTGGATC TTCTGAGGC CGATAACGGTC 6720  
 GTCTGCCCT CAAACTGGCA GATGCACGGT TACGATGCC CCATCTACAC CAACGTAACC 6780  
 TATCCCATT ACGTCAATCC GCCGTTTGTT CCCACGGAGA ATCCGACGGG TTGTTACTCG 6840  
 CTCACATTTA ATGTTGATGA AAGCTGGCTA CAGGAAGGCC AGACGCGAAT TATTTTGAT 6900  
 GCGGTTCCCTA TTGGTTAAAA AATGAGCTGA TTTAACAAAA ATTTAACGCC AATTTTAACA 6960  
 AAATATTAAC GTTACAATT TAAATATTTG CTTATACAAT CTTCTGTTT TTGGGGCTTT 7020  
 TCTGATTATC AACCGGGGTA CATATGATTG ACATGCTAGT TTTACGATTA CCGTTCATCG 7080  
 ATTCTCTTGT TTGCTCCAGA CTCTCAGGCA ATGACCTGAT AGCCTTGTGAT GATCTCTCAA 7140  
 AAATAGCTAC CCTCTCCGGC ATTAATTTAT CAGCTAGAAC GGTTGAATAT CATATTGATG 7200  
 GTGATTTCAC TGCTCCGGC CTTTCTCACC CTTTGAATC TTTACCTACA CATTACTCAG 7260  
 GCATTGCATT TAAAATATAT GAGGGTTCTA AAAATTTTA TCCTGCGTT GAAATAAAGG 7320  
 CTTCTCCGGC AAAAGTATTA CAGGGTCATA ATGTTTTGG TACAACCGAT TTAGCTTTAT 7380  
 GCTCTGAGGC TTATGCTT AATTTGCTA ATCTTIGCC TTGCTGTAT GATTTATTGG 7440  
 ACGTT 7445

## (2) INFORMATION FOR SEQ ID NO:2:

- (1) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 7317 base pairs
  - (B) TYPE: nucleic acid
  - (C) STRANDEDNESS: both
  - (D) TOPOLOGY: circular

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO:2:

AATGCTACTA CTATTAGTAG AATTGATGCC ACCTTTCAAG CTCGGCCCCC AAATGAAAAT 60  
 ATAGCTAAAC AGGTATTGAA CCATTTGCCA AATGTATCTA ATGGTAAAC TAAATCTACT 120  
 CGTTCCGAGA ATGGGAATC AACTGTTACA TGGAAATGAAA CTCCAGACA CCGTACTTTA 180  
 GTGCAATT TAAAACATGT TGAGCTACAG CACCAAGATTG AGCAATTAAAG CTCTAAGCCA 240  
 TCCGAAAAAA TGACCTCTTA TCAAAAGGAG CAATTAAAGG TACTCTCTAA TCCTGACCTG 300  
 TTGGAGTTG CTTCCGGTCT CGTTCCGTTT GAAGCTCGAA TTAAAACGGG ATATTTGAAG 360  
 TCTTTGGGC TTCTCTTAA TCTTTTGAT GCAATCCGCT TTGCTCTGA CTATAATAGT 420  
 CAGGGTAAAG ACCTGATTT TGATTTATGG TCATTCTCGT TTTCTGAACG GTTAAAGCA 480  
 TTTCAGGGGG ATTCAATGAA TATTTATGAC GATTCCGGCAG TATTGGACGC TATCCAGTCT 540  
 AACATTTTA CTATTACCCC CTCTGGAAA ACTTCTTTG CAAAAGCCTC TCGCTATTTT 600  
 GGTTTTATC GTCTGCTGGT AAACGAGGGT TATGATAGTC TTGCTCTTAC TATGCCTCGT 660  
 AATTCCTTTT GCGGTATGT ATCTGCATTA GTTGAATGTG GTATTCTAA ATCTCAACTG 720

ATGAATCTTT	CTACCTGTAA	TAATGTTGTT	CCGTTAGTTC	GTTTATTAA	CGTAGATTIT	780
TCTTCCAAC	GTCCGTACTG	GTATAATGAG	CCAGTTCTTA	AAATCGCATA	AGGTAATTCA	840
CAATGATTAA	AGTTGAAATT	AAACCATCTC	AAGCCCAATT	TACTACTCGT	TCTGGTGTIT	900
CTCGTCAGGG	CAAGCCTAT	TCACTGAATG	AGCAGCTTIG	TTACGTTGAT	TTGGGTAAATG	960
AATATCCGGT	TCTTGTCAAG	ATTACTCTTG	ATGAAGGTCA	GCCAGCCTAT	GCGCCTGGTC	1020
TGTACACCGT	TCATCTGTCC	TCTTICAAAG	TTGGTCAGTT	CGGTTCCCTT	ATGATTGA&C	1080
GTCTGCCCT	CGTICCGGGT	AAAGAACATG	GAGCAGGTCG	CGGATTTCGA	CACAATTAT	1140
CAGGCGATGA	TACAAATCTC	CGTTGTACTT	TGTTTCCGGC	TTGGTATAAT	CGCTGGGGT	1200
CAAAGATGAG	TGTTTGTAGTG	TATTCTTCG	CCTCTTCGT	TTAGGTTGG	TGCCTTCGTA	1260
GTGGCATTAC	GTATTTTACCG	CGTTTAATGG	AAACTTCCTC	ATGAAAAAGT	CTTTAGTCCT	1320
CAAAGCCTCT	GTAGCCGTG	CTACCTCGT	TCCGATGCTG	TCTTTCGCTG	CTGAGGGTGA	1380
CGATCCCGCA	AAAGCCGCCT	TTAACTCCCT	GCAAGCCTCA	GCGACCGAAT	ATATCGGTTA	1440
TGGTGGGGCG	ATGGTTGTG	TCATTGTCGG	CGCAACTATC	GGTATCAAGC	TGTTTAAGAA	1500
ATTCACCTCG	AAAGCAAGCT	GATAAACCGA	TACAATTAAA	GGCTCCTTTT	GGAGCCTTTT	1560
TTTTGGAGA	TTTCAACGT	GAAAAAATTAA	TTATTCGCAA	TTCCCTTGT	TGTTCCCTTC	1620
TATTCTCACT	CCGCTGAAAC	TGTTGAAAGT	TGTTTAGCAA	AACCCCATAAC	AGAAAATTCA	1680
TTTACTAACG	TCTGGAAAGA	CGACAAAATC	TTAGATCGTT	ACGCTAACTA	TGAGGGTTGT	1740
CTGTGGAATG	CTACAGGCCT	TGTAGTTGT	ACTGGTGACG	AAACTCAGTG	TTACGGTACA	1800
TGGGTTCTA	TTGGGCTTGC	TATCCCTGAA	AATGAGGGTG	GTGGCTCTGA	GGGTGGCGGT	1860
TCTGAGGGTG	GGGGTTCTGA	GGGTGGCGGT	ACTAAACCTC	CTGAGTACCG	TGATACACCT	1920
ATTCCGGCT	ATACTTATAT	CAACCCCTCTC	GACGGCACTT	ATCCCCCTGG	TACTGAGCAA	1980
AACCCCGCTA	ATCCTAATCC	TTCTCTTGAG	GAUTCTCAGC	CTCTTAATAC	TTTCATGTTT	2040
CAGAATAATA	GGTCCGAAA	TAGGCAGGGG	GCATTAACTG	TTTATACGGG	CACTGTTACT	2100
CAAGGCAGTG	ACCCCGTTAA	AACTTATTAC	CACTACACTC	CTGTATCATC	AAAAGCCATG	2160
TATGACGGTT	ACTGGAACGG	AAAATCAGA	GAUTGCGGT	TCCATTCTGG	CTTTAATGAA	2220
GATCCATTCTG	TTTGTGAATA	TCAAGGCCAA	TCGTCTGACC	TGCGCTAAC	TCCTGTCAAT	2280
GCTGGCGGGG	GCTCTGGTGG	TGGTTCTGGT	GGCGGCTCTG	AGGGTGGTGG	CTCTGAGGGT	2340
GGCGGTTCTG	AGGGTGGCGG	CTCTGAGGGG	GGCGGTTCCG	GTGGTGGCTC	TGTTCCGGT	2400
GATTTTGATT	ATGAAAAGAT	GGCAAACGCT	AATAAGGGGG	CTATGACCGA	AAATGCCGAT	2460
AAAAACGGCCG	TACAGTCTGA	CGCTAAAGGC	AAACTTGATT	CTGTCGCTAC	TGATTACGGT	2520
GCTGCTATCG	ATGGTTTCAT	TGGTGACGTT	TCCGGCTTIG	CTAATGGTAA	TGGTGCTACT	2580
GGTGATTITG	CTGGCTCTAA	TTCCCAAATG	GCTCAAGTCG	GTGACGGTGA	TAATTACACCT	2640
TTAATGAATA	ATTTCCGTCA	ATATTTACCT	TCCCTCCCTC	AATCCGGTGA	ATGTCGCCCT	2700
TTTGTCTTTA	GGCGCTGGTAA	ACCATATCAA	TTTCTATTG	ATTGTGACAA	AATAAACTTA	2760

TTCCGTGGTG TCTTTCGGTT TCTTTATAT GTGCCACCT TTATGTATGT ATTTTCTACG	2820
TTTGCTAAC A TACTGCGTAA TAAGGAGTCT TAATCATGCC AGTTCTTGT CGTATTCCGT	2880
TATTATTCGG TTTCCTCGGT TTCCCTCTGG TAACTTTGTT CGGCTATCTG CTTACTTTTC	2940
TTAAAAAGGG CTICGGTAAG ATAGCTATTG CTATTTCAATT CTTTCTTGCT CTTATTATTG	3000
GGCTTAACTC AATTCTGTG GGTTATCTCT CTGATATTAG CGCTCAATTAA CCCTCTGACT	3060
TTGTTCAAGG TGTTCAGTTA ATTCTCCCGT CTAATGCCGT TCCCTGTTT TATGTTATTC	3120
TCTCTGTAAA CGCTGCTATT TTCATTITG ACGTTAAACA AAAAATCGTT TCTTATTG	3180
ATTGGGATAA ATAATATGGC TGTTTATTGT GTAACTGGCA AATTAGGCTC TGGAAAGACG	3240
CTCGTTAGCG TTGGTAAGAT TCAGGATAAA ATTGTAGCTG GGTGCAAAT AGCAACTAAT	3300
CTTGATTAA GGCTTCAAAA CCTCCCGAA GTCCGGAGGT TCGCTAAAAC GCCTCGCGTT	3360
CTTACAATAC CCGATAAGCC TTCTATATCT GATTGCTTG CTATTGGCG CGGTAATGAT	3420
TCCTACGATG AAAATAAAA CGGCTTGCTT GTTCTCGATG AGTGGGTAC TTGGTTAAT	3480
ACCCGTTCTT GGAATGATAA GGAAAGACAG CCCATTATTG ATTGGTTCT ACATGCTCGT	3540
AAATTAGGAT GGGATATTAT TTTTCTGTGTT CAGGACTTAT CTATTGTTGA TAAACAGGGG	3600
CGTTCTGCAT TAGCTGAACA TCTTGTAT TGTCTCGTC TGGACAGAAT TACTTTACCT	3660
TTTGTGGTA CTTTATATTC TCTTATTACT GGCTCGAAAA TGCCTCTGCC TAAATTACAT	3720
GTTGGCGTTG TTAAATATGG CGATTCTCAA TIAAGCCCTA CTGTTGAGCG TTGGCTTAT	3780
ACTGGTAAGA ATTGTATAA CGCATATGAT ACTAACACAGG CTTTTCTAG TAATTATGAT	3840
TCCGGTGTGTT ATTCTTATTT AACGCCATTAT TTATCACACG GTCGGTATT CAAACCATTAA	3900
AATTAGGTC AGAACATGAA GCTTACTAAA ATATATTGAA AAAAGTTTTC ACGCGTTCTT	3960
TGTCTTGGGA TTGGATTTCGC ATCAGGATT ACATATAGTT ATATAACCCA ACCTAACCG	4020
GAGGTTAAAA AGGTAGTCTC TCAGACCTAT GATTTCTATA AATTCACTAT TGACTCTTCT	4080
CAGCGTCTTA ATCTAAGCTA TCGCTATGTT TTCAAGGATT CTAAGGAAA ATTAATTAAAT	4140
AGCGACGATT TACAGAAGCA AGGTTATTCA CTCACATATA TTGATTATG TACTGTTCC	4200
ATTAAAAAAAG GTAATTCAA TGAAATTGTT AAATGTAATT AATTITGTT TCTTGATGTT	4260
TGTTTCACTCA TCTTCTTTG CTCAGGTAAT TGAAATGAAT AATTCCGCTC TGCGCGATT	4320
TGTAACCTGG TATTCAAAGC AATCAGGCCA ATCCGTTATT GTTCTCCCG ATGAAAAGG	4380
TACTGTTACT GTATATTCACT CTGACGTTAA ACCTGAAAAT CTACGCAATT TCTTTATTC	4440
TGTTTACGT GCTAATAATT TTGATATGGT TGGTCAATT CCTTCCATAA TTCAAGAAGTA	4500
TAATCCAAAC AATCAGGATT ATATTGATGA ATTGCCATCA TCTGATAATC AGGAATATGA	4560
TGATAATTCC GCTCCTCTG GTGGTTCTT TGTTCCGAA AATGATAATG TTACTCAAAC	4620
TTTAAAAATT AATAACGTTG GGGCAAAGGA TTAAATACGA GTTGTGAAAT TGTTGTAAA	4680
GTCTAATACT TCTAAATCCT CAAATGTATT ATCTATTGAC CGCTCTAATC TATTAGTTGT	4740
TAGTGCACCT AAAGATATT TAGATAACCT TCCTCAATT CTTTCTACTG TTGATTG	4800

AACTGACCAAG ATATTGATTG AGGGTTTGAT ATTTGAGGTT CAGCAAGGTG ATGCTTTAGA 4860  
 TTTTCATTT GCTGCTGGCT CTCAGCGTGG CACTGTTGCA GGCGGTGTTA ATACTGACCG 4920  
 CCTCACCTCT GTTTTATCTT CTGCTGGTGG TTGCTTCGGT ATTTTTAATG GCGATGTTTT 4980  
 AGGGCTATCA GTTCGCCCAT TAAAGACTAA TAGCCATTCA AAAATATTGT CTGTGCCACG 5040  
 TATTCTTACG CTTTCAGGTC AGAAGGGTTC TATCTCTGTT GGCCAGAATG TCCCTTTTAT 5100  
 TACTGGTCGT GTGACTGGTG AATCTGCCAA TGTAATAAT CCATTTAGA CGATTGAGCG 5160  
 TCAAAATGTA GGTATTTCCA TGAGCGTTTT TCCTGTTGCA ATGGCTGGCG GTAATATTGT 5220  
 TCTGGATAATT ACCAGCAAGG CCGATAGTTT GAGTTCTTCT ACTCAGGCAA GTGATGTTAT 5280  
 TACTAATCAA AGAAGTATTG CTACAACGGT TAATTTGGT GATGGACAGA CTCTTTACT 5340  
 CGGTGGCCTC ACTGATTATA AAAACACTTC TCAAGATTCT GGCGTACCGT TCCTGTCTAA 5400  
 AATCCCCTTA ATCGGCCTCC TGTTTAGCTC CCGCTCTGAT TCCAACGAGG AAAGCACGTT 5460  
 ATACGTGCTC GTCAAAGCAA CCATAGTACG CGCCCTGTAG CGGCCCATTA AGCGCGCCGG 5520  
 GTGTGGTGGT TACGGCGAGC GTGACCGCTA CACTTGGCAG CGCCCTAGCG CCCGCTCCTT 5580  
 TCGCTTCTT CCCTTCCTT CTCGCCACGT TCGCCGGTT TCCCCGTCAA GCTCTAAATC 5640  
 GGGGGCTCCC TTIAGGGTTC CGATTIAGTC GTTACGGCA CCTCGACCCC AAAAAACTTG 5700  
 ATTTGGGTGA TGGTTCACGT AGTGGGCCAT CGCCCTGATA GACGGTTTTT CGCCCTTGA 5760  
 CGTTGGAGTC CACGTTCTT AATAGTGGAC TCTTGTICCA AACTGGAACA AACTCAACC 5820  
 CTATCTCGG CTATTCTTT GATTATAAG GGATTTGCC GATTTCGGAA CCACCATCAA 5880  
 ACAGGATTTT CGCCTGCTGG GGCAAACCGAG CGTGGACCCC TTGCTGCAAC TCTCTCAGGG 5940  
 CCAGGCGGTG AAGGGCAATC AGCTGTTGCC CGTCTCGCTG GTGAAAAGAA AAACCACCC 6000  
 GGCCCCAAT ACGCAAACCG CCTCTCCCCG CGCGTTGGCC GATTCAATTAA TGCAGCTGGC 6060  
 ACGACAGGTT TCCCGACTGG AAAGCGGGCA GTGAGCGCAA CGCAATTAAAT GTGAGTTAGC 6120  
 TCACTCAATT GGCAACCCAG CCTTACACT TTATGCTTCC GGCTCGTATG TTGTTGCAA 6180  
 TTGTGAGCGG ATAACAATT T CACACGCCAA GGAGACAGTC ATAATGAAAT ACCTATTGCC 6240  
 TACGGCAGCC GCTGGATTGT TATTACTCGC TGCCCAACCA GCCATGGCCG AGCTCGTGAT 6300  
 GACCCAGACT CCAGATATCC AACAGGAATG AGTGTAAATT CTAGAACCGC TCACTTGGCA 6360  
 CTGGCCGTCC TTTACAACG TCGTGAATGG GAAAACCTG CGTACCCCA AGCTTAATCG 6420  
 CCTTGGAGAA TTCCCTTCG CCAGCTGGCG TAATAGCGAA GAGGCCCGCA CCGATCGCCC 6480  
 TTCCCAACAG TTGGCAGCC TGAATGGCGA ATGGCGTTT GCCTGGTTTC CGGCACCGAGA 6540  
 AGCGGTGCCG GAAAGCTGGC TGGAGTGCAG TCTTCCGAG GCCGATACGG TCGTCGTCCC 6600  
 CTCAAACCTGG CAGATGCAGG GTTACGATGC GCCCATCTAC ACCAACGTAA CCTATCCCAT 6660  
 TACGGTCAAT CGCCCGTTT G TCCCCACCGA GAATCCGAGG GGTTGTTACT CGTCACATT 6720  
 TAATGTTGAT GAAAGCTGGC TACAGGAAGG CCAGACGCCA ATTATTTTG ATGGCGTTCC 6780  
 TATTGGTTAA AAAATCAGCT GATTTAACAA AAATTTAACCG CGAATTAA CAAAATATTA 6840

ACGTTTACAA TTAAATATT TGCTTATACA ATCTTCGTGT TTTGGGGCT TTTCTGATTA  
 TCAACCGGGG TACATATGAT TGACATGCTA GTTTACGAT TACCGTTCAT CGATTCTCTT  
 GTTGCCTCCA GACTCTCAGG CAATGACCTG ATAGCCTTG TAGATCTCTC AAAAATAGCT  
 ACCCTCTCCG GCATTAATT ATCAGCTAGA ACGGTTGAAT ATCATATTGA TGGTGATTIG  
 ACTGTCTCCG GCCTTCTCA CCCTTITGAA TCTTACCTA CACATTACTC AGGCATTGCA  
 TTTAAAATAT ATGAGGGTTC TAAAAATTAT TATCCTTGCG TTGAAATAAA GGCTTCTCCC  
 GCAAAAGTAT TACAGGGTCA TAATGTTTT GGTACAACCG ATTTAGCTTT ATGCTCTGAG  
 GCTTTATTGC TIAATTTCG TAATTCTTIG CCTTGCCTGT ATGATTATT GGATGTT  
 (2) INFORMATION FOR SEQ ID NO:3:

- (1) SEQUENCE CHARACTERISTICS:  
 (A) LENGTH: 7729 base pairs  
 (B) TYPE: nucleic acid  
 (C) STRANDEDNESS: both  
 (D) TOPOLOGY: circular

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO:3:

AATGCTACTA CTATTAGTAG AATTGATGCC ACCTTTTCAG CTCGGCCCC AAATGAAAAT  
 ATAGCTAAC AGGTATTGA CCATTGCGA AATGTATCTA ATGGTCAAAC TAAATCTACT  
 CGTTCGCAGA ATTGGAATC AACTGTTACA TGAATGAAA CTTCCAGACA CCGTACTTTA  
 GTTGCATATT TAAAACATGT TGAGCTACAG CACCAGATTG AGCAATTAAG CTCTAAGCCA  
 TCTGCAAAAA TGACCTCTTA TCAAAAGGAG CAATTAAAGG TACTCTCTAA TCCTGACCTG  
 TTGGAGTTG CTTCCGGTCT CGTTCGCTT GAAGCTCGAA TTAAAACCGG ATATTTGAAG  
 TCTTTCGGGC TTCTCTTAA TCTTTTGAT GCAATCCGCT TTGCTCTGA CTATAATAGT  
 CAGGGTAAAG ACCTGATTIT TGATTTATGG TCATTCTCGT TTTCTGAACG GTTTAAAGCA  
 TTGAGGGGG ATTCAATGAA TATTTATGAC GATTCCGGAG TATTGGACCC TATCCAGCT  
 AAACATTTA CTATTACCCC CTCTGGCAAA ACTTCTTTG CAAAAGCCTC TCGCTATT  
 GGTTTTTATC GTCGTCTGGT AAACGAGGGT TATGATAGTG TTGCTCTTAC TATGCCCTG  
 AATCCTTTT GGCGTTATGT ATCTGCATTA GTTGAATGTG GTATTCCTAA ATCTCAACTG  
 ATGAATCTTT CTACCTGTAA TAATGTTGTT CCGTTAGTTC GTTTTATTAA CGTAGATT  
 TCTTCCCAAC GTCCGTACTG GTATAATGAG CCAGTTCTTA AAATCCATA AGGTAAATTCA  
 CAATGATTAA AGTTGAAATT AAACCATCTC AAGCCCAATT TACTACTCGT TCTGGTGT  
 CTGGTCAAGGG CAAGCCTTAT TCACTGAATG AGCAGCTTG TTACGTTGAT TTGGGTAATG  
 AATATCCGGT TCTTGTCAAG ATTACTCTTG ATGAAGGTCA GCCAGCCTAT GCGCCTGGTC  
 TGTACACCGT TCATCTGTCC TCTTCAAAAG TTGGTCAGTT CGGTTCCCTT ATGATTGACC  
 GTCTGCGCCT CGTCCGGCT AAGTAACATG GAGCAGGTGCG CGGATTTCGA CACAATT  
 CAGGGATGA TACAAATCTC CGTTGTACTT TGTTCGGCC TTGGTATAAT CGCTGGGGGT  
 6900  
 6960  
 7020  
 7080  
 7140  
 7200  
 7260  
 7317

CAAAGATGAG TGTGTTAGTG TATTCTTCG CCTCTTICGT TTAGGTTGG TGCCCTCGTA	1260
GTGGCATTAC STATTTTACC CGTTTAATGG AAACCTTCCTC ATGAAAAAAGT CTITAGTCCT	1320
CAAAGCCTCT STAGCCGTG CTACCCCTCGT TCCGATGCTG TCITTCGCTG CTGAGGGTGA	1380
CGATCCCGCA AAAGCGGCCT TTAACTCCCT GCAAGCCTCA GCGACCGAAT ATATCGGTTA	1440
TCCGTGGCG AIGGTTGTG TCATTGTCGG CGCAACTATC GGTATCAAGC TGTTTAAGAA	1500
ATTCACCTCG AAAGCAAGCT GATAAACCGA TACAATTAAA GGCTCCCTTT GGAGCCTTT	1560
TTTTGGAGA TTTCACCGT GAAAAAAATTA TTATTGCAA TTCCCTTGT AGTTCCTTTC	1620
TATTCTCACT CCGCTGAAAC TGTTGAAAGT TGTTTAGCAA AACCCCATAAC AGAAAATTCA	1680
TTTACTAACG TCTGGAAAGA CGACAAAAGT TTAGATCGTT ACGCTAACTA TGAGGGTTGT	1740
CTGTGGAATG CTACAGGGT TGTAGTTGT ACTGGTGACG AAACTCAGT TTACGGTACA	1800
TGGGTTCTA TTGGGCTTGC TATCCCTGAA AATGACGGTG GTGGCTCTGA GGGTGGCGGT	1860
TCTGAGGGTG GCGGTTCTGA GGGTGGCGGT ACTAAACCTC CTGAGTACGG TGATACACCT	1920
ATCCCCGGCT ATACTTATAT CAACCCTCTC GACGGCACTT ATCCGCCCTGG TACTGAGCAA	1980
AACCCCGCTA ATCCTAATCC TTCTCTTGAG GAGTCTCAGC CTCTTAATAC TTTCATGTTT	2040
CAGAATAATA CGTTCCGAAA TAGGCAGGGG GCATTAACCTG TTATACGGG CACTGTTACT	2100
CAAGGCACCTG ACCCCGTTAA AACTTATTAC CAGTACACTC CTGTATCATC AAAAGCCATG	2160
TATGACCGTT ACTGGAACGG TAAATTAGA GACTGCGTT TCCATTCTGG CTTTAATGAA	2220
GATCCATTG TTTGTGAATA TCAAGGCCAA TCGTCTGACC TCCCTCAACC TCCGTCAAT	2280
GCTGGGGCG GCTCTGGTGG TGGTTCTGGT GCGGGCTCTG AGGGTGGTGG CTCTGAGGGT	2340
GGCGGTTCTG AGGGTGGCGG CTCTGAGGGA GCGGGTTCCG GTGGTGGCTC TGGTCCGGT	2400
GATTTGATT ATGAAAAGAT GGCAAAAGCT AATAAGGGGG CTATGACCGA AAATGCCAT	2460
GAAAAGCCGC TACAGTCTGA CGCTAAAGGC AAACCTGATT CTGTCGCTAC TGATTACGGT	2520
GCTGCTATEG ATGGTTTCAT TGGTACGGTT TCCGGCCTTG CTAATGGTAA TGGTGTACT	2580
GGTGATTITG CTGGCTCTAA TTCCCAAATG GCTCAAGTCC GTGACGGTGA TAATTACCT	2640
TTAATGAATA ATTCCGTCA ATATTACCT TCCCTCCCTC AATCGGTGA ATGTCGCCCT	2700
TTTGTCTTTA GCGCTGGTAA ACCATATGAA TTTTCTATTG ATTGTGACAA AATAAAACTTA	2760
TTCCGTGGTG TCTTTCGTT TCTTTATAT GTTGCCACCT TTATGTATGT ATTTCCTACG	2820
TTTGCTAACCA TACTGCGTAA TAAGGAGTCT TAATCATGCC AGTTCTTGTG GGTATTCCGT	2880
TATTATTGCG TTCCCTCGGT TTCCCTCTGG TAACTTTGTG CGGCTAICTG CTTACTTTTC	2940
TTAAAAAGGG CTTCCGTAAG ATAGCTATG CTATTCAITT GTTCTTGCT CTTATTATTG	3000
GGCTTAACTC AATTCTTG TGTTATCTCT CTGATATTAG CGCTCAAITA CCCTCTGACT	3060
TTGTTCAAGGG TGTTCAAGTAA ATTCTCCCGT CTAATCGGCT TCCCTGTTT TATGTTATTG	3120
TCTCTGTAAA GGCTGCTATT TTCAATTITG ACGTTAAACA AAAAATCGTT TCTTATTGG	3180
ATTGGGATAAA ATAATATGGC TGTTTATTTT GTAACGGCA AATTAGGCTC TGGAAAGACG	3240

CTCGTTAGCG TTGGTAAGAT TCAGGATAAA ATTGTAGCTG GGTGCAAAAT AGCAACTAAT 3300  
 CTTGATTTAA GGCTTCAAAA CCTCCCGCAA GTCGGGAGGT TCGCTAAAAC GCCTCGCGTT 3360  
 CTTAGAATAC CGATAAGCC TTCTATATCT GATTGCTTG CTATTGGCG CGGTAATGAT 3420  
 TCCTACGATG AAAATAAAAA CGGCTTGCTT GTTCTCGATG AGTGGGGTAC TTGGTTTAAT 3480  
 ACCCGTTCTT GGAATGATAA GGAAAGACAG CCGATTATTG ATTGGTTCT ACATGCTCGT 3540  
 AAATTAGGAT GGGATATTAT TTTTCTTGTGTT CAGGACTTAT CTATTGTTGA TAAACAGGCC 3600  
 CGTTCTGCAT TAGCTGAACA TGTTGTTAT TGTCGTCGTC TGGACAGAAT TACTTTACCT 3660  
 TTTGTGGTA CTTTATATTG TCTTATTACT GGCTCGAAAA TGCCTCTGCC TAAATTACAT 3720  
 GTTGGCGTTG TAAATATGG CGATTCTCAA TTAAGCCCTA CTGTTGAGCG TTGGCTTAT 3780  
 ACTGGTAAGA ATTGTATAA CGCATATGAT ACTAAACAGG CTTTTCTAG TAATTATGAT 3840  
 TCGGGTGTGTT ATTCTTATT AACGCCATTAT TTATCACACCG GTCGGTATT CAAACCATT 3900  
 AAITTAGGTC AGAAGATGAA GCTTACTAAA ATATATTGAA AAAAGTTTTC ACGCGTTCTT 3960  
 TGTCTTGGCA TTGGATTTCG ATCAGCATT ACATATAGTT ATATAACCCA ACCTAACGCCG 4020  
 GAGGTTAAAAA AGGTAGTCTC TCAGACCTAT GATTTGATA AATTCACTAT TGACTCTCT 4080  
 CAGCGTCTTA ATCTAAGCTA TCGCTATGTT TTCAAGGATT CTAAGGGAAA ATTAATTAAT 4140  
 AGCGACGATT TACAGAAGCA AGGTTATTCA CTCACATATA TTGATTITATG TACTGTTCC 4200  
 ATTAAAAAAAG GTAATTCAA TGAAATTGTT AAATGTAATT AATTITGTTT TCTTGATGTT 4260  
 TGTITCATCA TCTTCTTITG CTCAGGTAAT TGAAATGAAT AATTGCCCTC TGCGCGATT 4320  
 TGTAACTTGG TATTCAAAGC AATCAGGCCA ATCCGTTATT GTTICTCCCG ATGTAAGG 4380  
 TACTGTTACT GTATATTCACT CTGACGTTAA ACCTGAAAAT CTACGCAATT TCTTTATTTC 4440  
 TGTTTACGT GCTAATAATT TTGATATGGT TGGTCAATT CCTTCCATAA TTCAAGAGTA 4500  
 TAATCCAAAC AATCAGGATT ATATTGATGA ATTGCCATCA TCTGATAATC AGGAATATGA 4560  
 TGATAATTCA GCTCCTCTG GTGGTTCTT TGTTCCGCAA AATGATAATG TTACTCAAAC 4620  
 TTTTAAAATT AATAACGTTG GGGCAAAGGA TTTAATACGA GTTGTGAAAT TGTTTGTAAA 4680  
 GTCTAATACT TCTAAATCCT CAAATGTATT ATCTATTGAC GGCTCTAAATC TATTAGTTGT 4740  
 TAGTGCACCT AAAGATATT TAGATAACCT TCCTCAATTCT TTTTCTACTG TTGATTGCCC 4800  
 AACTGACCAAG ATATTGATTG AGGGTTGAT ATTGAGGTT CAGCAAGGTG ATGCTTTAGA 4860  
 TTTTCAATT GCTGCTGGCT CTCAGCGTGG CACTGTTGCA GGCGGTGTTA ATACTGACCG 4920  
 CCTCACCTCT GTTTATCTT CTGCTGGTGG TTGGTTGGT ATTGTTAATG GCGATGTTTT 4980  
 AGGGCTATCA GTTCCGGCAT TAAAGACTAA TAGCCATTCA AAAATATTGT CTGTGCCACCG 5040  
 TATTCTTACG CTTTCAGGTC AGAAGGGTTC TATCTCTGTT GGCCAGAATG TCCCTTTAT 5100  
 TACTGGTGTGTT GTGACTGGTG AATCTGCCAA TGAAATAAT CCATTTGAGA CGATTGAGCC 5160  
 TCAAAATGTA GGTATTCCA TGAGCGTTT TCCCTGTTGCA ATGGCTGGCG GTAATATTGT 5220  
 TCTGGATATT ACCACCAAGG CCGATAGTTT GAGTTCTTCT ACTCAGGCCAA GTGATGTTAT 5280

TACTAATCAA AGAAAGTATTG CTACAACGGT TAATTTGGT GATGGACAGA CTCTTTACT 5340  
 CGGTGGCCTC ACTGATTATA AAAACACTTC TCAAGATTCT GGCGTACCGT TCCTGCTAA 5400  
 AATCCCTTA ATCGGCCTCC TGTTTAGCTC CCGCTCTGAT TCCAACGAGG AAAGCACGTT 5460  
 ATACGTGCTC GTCAAAGCAA CCATAGTAGG CCCCTGTAG CGGGCGATTA AGGCCGGCGG 5520  
 GTGTGGTGGT TACGCGCAGC GTGACCGCTA CACTTGCCAG CGCCCTAGCG CCCGCTCCTT 5580  
 TCGCTTCTT CCCTTCCTT CTCGCCACGT TCGCCGGCTT TCCCCGTCAA GCTCTAAATC 5640  
 GGGGGCTCCC TTAGGGTTC CGATTTAGTG CTTTACGGCA CCTCGACCCC AAAAAACTTG 5700  
 ATTTGGTGA TGTTTCACGT AGTGGGCCAT CGCCCTGATA GACGGTTTT CGCCCTTGA 5760  
 CGTTGGAGTC CACGTTCTT AATAGTGGAC TCTTGTCCA AACTGGAACA ACACCTAAC 5820  
 CTATCTCGGG CTATTCTTT GATTATAAG GGATTTGCC GATTTCGGAA CCACCATCAA 5880  
 ACAGGATTT CCCCTGCTGG GCAAACCAAG CGTGGACCCG TTGCTGCAAC TCTCTCAGGG 5940  
 CCAGGGCGTG AAGGCAATC AGCTGTTGCC CGTCTCGCTG GTGAAAAGAA AAACCACCC 6000  
 GGCCCCAAT ACGCAAACCG CCTCTCCCCG CGCGTTGGCC GATTCAATTAA TGCAGCTGGC 6060  
 ACGACAGGTT TCCCGACTGG AAAGCGGGCA GTGAGCGCAA CGCAATTAAAT GTGAGTTAGC 6120  
 TCACTCAITTA GGCAACCCAG GCTTIACT TTATGCTTCC GGCTCGTATG TTGTGTGGAA 6180  
 TTGTGAGCGG ATAACAATT CACACGGTC ACTTGGCACT GGCCGTCGTT TTACAACGTC 6240  
 GTGACTGGGA AAACCTGGC GTTACCCAAG CTTGTACAT GGAGAAAATA AAGTGAAC 6300  
 AAGCACTATT GCACCTGGCAC TCTTACCGTT ACTGTTTACC CCTGTGGCAA AAGCCCAGGT 6360  
 CCAGCTGCTC GAGTCGGTCT TCCCCCTGGC ACCCTCCTCC AAGAGCACCT CTGGGGCAC 6420  
 AGCGGCCCTG GGCTGCCTGG TCAAGACTAA TTCCCCGAAC CGGTGACGGT GTCGTGGAAC 6480  
 TCAGGGCGCC TGACCAGCGG CGTGCACACC TTCCCCGGCTG TCCTACAGTC CTCAGGACTC 6540  
 TACTCCCTCA GCAGCGTGGT GACCGTGGCC TCCAGGAGCT TGGGCACCCA GACCTACATC 6600  
 TCCAACGTCA ATCACAAAGCC CAGCAACACC AAGGTGGACA AGAAAGCAGA GCCCAAATCT 6660  
 TGTACTAGTG GATCCTACCC GTACGACGTT CCGGACTACG CTTCTTAGGC TGAAGGGCAT 6720  
 GACCCGTCA AGGCTGCATT CAATAGTTA CAGGCAAGTG CTACTGAGTA CATTGGCTAC 6780  
 GCTTGGGCTA TGGTAGTAGT TATAGTTGGT CCTACCATAG GGATTAAATT ATTCAAAAG 6840  
 TTTACGAGCA AGGCTTCTTA AGCAATACCG AAGAGGCCCG CACCGATCGC CCTTCCCAAC 6900  
 AGTTGGCGAG CCTGAATGGC GAATGGCGCT TTGCTGGTT TCCGGCACCA GAAGCGGTGC 6960  
 CGGAAAGCTG GCTGGAGTGC GATCTTCCCTG AGGCCGATAC GGTCGTGTC CCCTCAAAC 7020  
 GGCAGATGCA CGGTACCGAT CGGCCCATCT ACACCAACGT AACCTATCCC ATTACGGTCA 7080  
 ATCCGCCGT TGTTCGGACG GAGAATCCGA CGGGTTGTTA CTGGCTCACA TTTAATGTTG 7140  
 ATGAAAGCTG GCTACAGGAA GGCCAGACGC GAATTATTT TGATGGCGTT CCTATTGGTT 7200  
 AAAAAATGAG CTGATTTAAC AAAAATTAA CGCGAATTAA AACAAAATAT TAACGTTTAC 7260  
 AATTTAAATA TTGCTTATA CAATCTTCCG GTTTTGGGG CTTTCTGAT TATCAACCGG 7320

GGTACATATG ATTGACATGC TAGTTTACG ATTACCGTTC ATCGATTCTC TTGTTGCTC 7380  
 CAGACTCTCA GGCAATGACC TGATAGCCTT TGTAGATCTC TCAAAAATAG CTACCCCTCTC 7440  
 CGGCATTAAT TTATCAGCTA GAACGGTTGA ATATCATATT GATGGTGATT TGACTGTCTC 7500  
 CGGCCTTTCT CACCCCTTTG AATCTTAC TACACATTAC TCAGGCATTG CATTAAAAT 7560  
 ATATGAGGGT TCTAAAAATT TTTATCCTTG CGTTGAAATA AAGGCTTCTC CCGCAAAAGT 7620  
 ATTACAGGGT CATAATGTTT TTGGTACAAC CGATTTAGCT TTATGCTCTG AGGCTTTATT 7680  
 GCTTAATTTC GCTAATTCTT TGCCTTGCCT GTATGATTIA TTGGACGTT 7729

## (2) INFORMATION FOR SEQ ID NO:4:

- (i) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 7557 base pairs
  - (B) TYPE: nucleic acid
  - (C) STRANDEDNESS: both
  - (D) TOPOLOGY: circular

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO:4:

AATGCTACTA CTATTAGTAG AATTGATGCC ACCTTTTCAAG CTGGCGCCCC AAATGAAAAT 60  
 ATAGCTAAC AGGTATTGAG CCATTGCGA AATGTATCTA ATGGTCAAAC TAAATCTACT 120  
 CGTTCCAGA ATTGGGAATC AACTGTTACA TGGAATGAAA CTTCCAGACA CCGTACTTTA 180  
 GTTGCATATT TAAAACATGT TGAGCTACAG CACCAAGATTC AGCAATTAAAG CTCTAAGCCA 240  
 TCCGCAAAAA TGACCTCTTA TCAAAAGGAG CAATTAAAGG TACTCTCTAA TCCTGACCTG 300  
 TTGGAGTTTG CTCCGGTCT GGTTGCTTT GAAGCTCGAA TTAAAACCCG ATATTGAAAG 360  
 TCTTCGGGC TTCCCTTTAA TCTTTTGAT GCAATCCGCT TTGCTTCTGA CTATAATAGT 420  
 CAGGGTAAAG ACCTGATTIT TGATTATGG TCATTCTCGT TTTCTGAACG GTTTAAAGCA 480  
 TTTGAGGGGG ATTCAATGAA TATTTATGAC GATTCCCGAG TATTGGACGC TATCCAGTCT 540  
 AAACATTITA CTATTACCCC CTCTGCCAAA ACTTCTTTTG CAAAAGCCTC TCGCTATTIT 600  
 GGTTTTATC GTCGTCTGGT AAACGAGGGT TATGATAGTG TTGCTCTTAC TATGCCCTCGT 660  
 AATTCCCTTT GGCGTTATGT ATCTGCATTA GTTGAATGTG GTATTCTAA ATCTCAACTG 720  
 ATGAATCTTT CTACCTGTAA TAATGTTGTT CCGTTAGTTC GTTTTATTAA CGTAGATTIT 780  
 TCTTCCCAAC GTCCGTACTG GTATAATGAG CCAGTTCTTA AAATCCGATA AGGTAATTCA 840  
 CAATGATTAA AGTGAATT AAACCATCTC AAGCCCAATT TACTACTCGT TCTGGTGTIT 900  
 CTCGTCAGGG CAAGCCTTAT TCACTGAATG AGCAGCTTIG TTACGTTGAT TTGGGTAATG 960  
 AATATCCGGT TCTTGTCAAG ATTACTCTG ATGAAGGTCA GCCAGCCTAT GCGCCTGGTC 1020  
 TGTACACCGT TCATCTGTCC TCTTCAAAAG TTGGTCAGTT CGGTTCCCTT ATGATTGACC 1080  
 GTCTGCCCT CGTTCGGCT AAGTAACATG GAGCAGGTGCG CGGATTTCGA CACAATTAT 1140  
 CAGGCCATCA TACAAATCTC CGTTGTACTT TGTTCGCCC TTGGTATAAT CCCTGGGGGT 1200  
 CAAAGATGAG TGTTTGTG TATTCTTCTG CCTCTTCTGT TTTAGGTTGG TGCCTTCTGTA 1260

GTGGCATTAC GTATTTTACCGTTTAATGG AAACCTTCCTC ATGAAAAAGT CTTTAGTCCT 1320  
 CAAAGCCTCT GTAGCCGTG CTACCCCTCGT TCCGATGCTG TCTTTCGCTG CTGAGGGTGA 1380  
 CGATCCCGCA AAAGCGGCCT TTAACTCCCT GCAAGCCTCA GCGACCGAAT ATATCGGTTA 1440  
 TCGCTGGCG ATGGTTGTG TCATTGTCGG CGCAACTATC GGTATCAAGC TGTTTAAGAA 1500  
 ATTCACTCG AAAGCAAGCT GATAAACCGA TACAATTAAA GGCTCCTTT GGAGCCTTT 1560  
 TTTTGGAGA TTTTCAACGT GAAAAAAATTA TTATTCGAA TTCCCTTAGT TGTTCCCTTC 1620  
 TATTCTCACT CGCGTGAAC TGTTGAAAGT TGTTTAGCAA AACCCCATAAC AGAAAATTC 1680  
 TTTACTAACG TCTGGAAAGA CGACAAAATC TTAGATCGTT ACGCTAACTA TGAGGGTTGT 1740  
 CTGTGGAATG CTACAGGCGT TGTAGTTGT ACTGGTGACG AAACTCAGTG TTACGGTACA 1800  
 TGGGTTCTA TTGGGCTTGC TATCCCTGAA AATGAGGGTG GTGGCTCTGA GGCTGGCGGT 1860  
 TCTGAGGGTG CGCGTTCTGA GGGTGGCGGT ACTAAACCTC CTGAGTACGG TGATACACCT 1920  
 ATTCCGGGCT ATACTTATAT CAACCCTCTC GACGGCACTT ATCCGGCTGG TACTGAGCAA 1980  
 AACCCCGCTA ATCCTAATCC TTCTCTTGAG GAGTCTCAGC CTCTTAATAC TTTCATGTT 2040  
 CAGAATAATA GGTTCGAAA TAGGCAGGGG GCATTAAC TG TTTATACGGG CACTGTTACT 2100  
 CAAGGCAC TG ACCCCGTTAA AACTTATTAC CAGTACACTC CTGTATCATC AAAAGCCATG 2160  
 TATGACCGCTT ACTCGGAACGG TAAATTAGA GACTGCGCTT TCCATTCTGG CTTTAATGAA 2220  
 GATCCATTG TTTGTGAATA TCAAGGCCAA TCGTCTGACC TGCCTCAACC TCCGTCAAT 2280  
 GCTGGCGGGG GCTCTGGTGG TGGTTCTGGT GGCGGCTCTG AGGGTGGTGG CTCTGAGGGT 2340  
 GGCGGTTCTG AGGGTGGCGG CTCTGAGGGA GCGGGTTCCG GTGGTGGCTC TGGTTCCGGT 2400  
 GATTTGATT ATGAAAAGAT GGCACACGCT AATAAGGGGG CTATGACCGA AAATGCCGAT 2460  
 GAAAACGGGC TACAGTCTGA CGCTAAAGGC AAACCTGATT CTGTGCTAC TGATTACGGT 2520  
 GCTGCTATCG ATGGTTTCAT TGGTGACGTT TCCGGCCTTG CTAATGGTAA TGGTGCTACT 2580  
 GGTGATTTC CTGGCTCTAA TTCCCAAATG GCTCAAGTCG GTGACGGTGA TAATTCACCT 2640  
 TTAATGAATA ATTTCCGTCA ATATTTACCT TCCCTCCCTC AATCGGTGAA ATGTCGCCCT 2700  
 TTGTCCTTAA GCGCTGGTAA ACCATATGAA TTTCTATTG ATGTCGACAA AATAAAACTTA 2760  
 TTCCGTGGTG TCTTTCGCTT TCTTTTATAT GTGCCACCT TTATGTAATG ATTTCTACG 2820  
 TTGCTAACG TACTGGTAA TAAGGAGTCT TAATCATGCC AGTTCTTTTG GGTATTCCGT 2880  
 TATTATTGGG TTCCCTCGGT TTCCCTCTGG TAACTTTGTT CGGCTATCTG CTACTTTTC 2940  
 TTAAAAAGGG CTTGGTAAAG ATAGCTATTG CCTGTTCTT GCTCTTATTA TTGGGCTTAA 3000  
 CTCAATTCTT GTGGGTIATC TCTCTGATAT TAGCGCTCAA TTACCCCTCTG ACTTTGTTCA 3060  
 GGGTGTTCAG TTAATTCTCC CGTCTAAATGC GCTCCCTGT TTTTATGTTA TTCTCTCTGT 3120  
 AAAGGCTGCT ATTTCAATT TTGACGTTAA ACAAAAAATC GTTCTTATT TGGATTGGGA 3180  
 TAAATAATAT GGCTGTTTAT TTGTAAC TG GCAAATTAGG CTCTGGAAAG ACGCTCGTTA 3240  
 GCCTTGGTAA GATTCAAGGAT AAAATTGTAG CTGGGTGCAA AATAGCAACT AATCTTGATT 3300

TAAGGCTTCA AAACCTCCCC CAAGTCGGGA GGTTCGCTAA AACGCCCTCGC GTTCTTAGAA 3360  
 TACCGGATAA CCCTTCTATA TCTGATTTGC TTGCTATTGG GCGCGGTAAAT GATTCCCTACG 3420  
 ATGAAAATAA AAACGGCTTG CTTGTTCTCG ATGAGTCCGG TACTTCGGTT AATACCCGTT 3480  
 CTTGGAATGA TAAGGAAAGA CAGCCGATTA TTGATTGGTT TCTACATGCT CGTAAATTAG 3540  
 GATGGGATAT TATTTTCTT GTTCAGGACT TATCTATTGT TGATAAACAG GCGCGTTCTG 3600  
 CATTAGCTGA ACATGTTGTT TATTGTCGTC GTCTGGACAG AATTACTTTA CCTTTTGTGG 3660  
 GTACTTTATA TTCTCTTATT ACTGGCTCGA AAATGCCCTCT GCCTAAATTAA CATGTTGGCG 3720  
 TTGTTAAATA TGGCGATTCT CAATTAAGCC CTACTGTTGA GCGTTGGCTT TATACTGGTA 3780  
 AGAATTGTA TAACGGCATAT GATACTAAAC AGGCTTTTC TAGTAATTAT GATTCCGGTG 3840  
 TTTATTCTTA TTTAACGGCT TATTTATCAC ACGGTCCGGTA TTTCAAACCA TTAAATTAG 3900  
 GTCAGAAGAT GAAGCTTACT AAAATATAATT TGAAAAAGTT TTCACGGCTT CTTTGTCTTG 3960  
 CGATTGGATT TCCATCAGCA TTTACATATA GTTATATAAC CCAACCTAAG CCCGAGGGTA 4020  
 AAAAGGTAGT CTCTCAGACC TATGATTTG ATAAATTACAC TATTGACTCT TCTCAGCGTC 4080  
 TTAATCTAAG CTATCGCTAT GTTTCAAGG ATTCTAAGGG AAAATTAAATT AATAGCGACG 4140  
 ATTTACAGAA GCAAGGTTAT TCACTCACAT ATATTGATTT ATGTAUTGTT TCCATTAAAA 4200  
 AAGGTAATTG AAATGAAATT GTTAAATGTA ATTAATTGTT TTTTCTTGAT GTTTGTITCA 4260  
 TCATCTTCTT TTGCTCAGGT AATTGAAATG AATAATTCCG CTCTCGCGCA TTTTGTAACT 4320  
 TGGTATTCAA AGCAATCAGG CGAATCCGTT ATTGTTCTC CCGATGTAAA AGGTACTGTT 4380  
 ACTGTATATT CATCTGACGT TAAACCTGAA AATCTACGCCA ATTTCTTAT TTCTGTTTA 4440  
 CGTGCTAATA ATTTGATAT CGTTGGTTCA ATTCCCTCCA TAATTAGAA GTATAATCCA 4500  
 AACAAATCAGG ATTATAATTGA TGAATTGCCA TCATCTGATA ATCAGGAATA TGATGATAAT 4560  
 TCCGCTCCTT CTGGTGGTTT CTTTGTCCG CAAAATGATA ATGTTACTCA AACTTTAAA 4620  
 ATTAATAA&G TTCCGGCAAA CGATTTAATA CGAGTTGTCG AATTGTTTGT AAAGTCTAAT 4680  
 ACTTCTAAAT CCTCAAATGT ATTAATCTATT GACGGCTCTA ATCTATTAGT TGTTAGTGCA 4740  
 CCTAAAGATA TTTAGATAA CCTTCCTCAA TTCCCTTCTA CTGTTGATTT CCCAACTGAC 4800  
 CAGATATTGA TTGAGGGTTT GATATTGAG GTTCAGGAAG GTGATGCTTT AGATTTTCA 4860  
 TTGCTGCTG GCTCTCAGCG TGGCACTGTT GCAGGGCGGTG TTAATACTGA CCGCCTCACC 4920  
 TCTGTTTAT CTCTGCTGG TGGTTCGTTC GGTATTTTA ATGGCGATGT TTIAGGGCTA 4980  
 TCAGTTCCGG CATTAAAGAC TAATAGCCAT TCAAAAATAT TGTCTGTGCC ACGTATTCTT 5040  
 ACGCTTTCAG GTCAGAAGGG TTCTATCTCT GTTGGCCAGA ATGTCCTTT TATTACTGGT 5100  
 CGTGTGACTG GTGAATCTGC CAATGTAAT AATCCATTTC AGACGATTGA CGGTCAAAAT 5160  
 GTAGGTATT CCATGAGCGT TTTCTCTGTT GCAATGGCTG GCGGTAAATAT TGTCTGGAT 5220  
 ATTACCAAGCA AGGCCGATAG TTTCAGTTCT TCTACTCAGG CAAGTGATGT TATTACTAAT 5280  
 CAAAGAAGTA TICCTACAAC CGTTAATTG CGTCACTGGAC AGACTCTTT ACTCGGTGGC 5340

CTCACTGATT ATAAAACAC TTCTCAAGAT TCTGGCGTAC CGTTCCGTGTC TAAAATCCCT 5400  
 TTAATCGGCC TCCTGTTAG CTCCCGCTCT GATTCCAACG AGGAAAGCAG GTTATACTGC 5460  
 CTCGTCAAAG CAACCATACT ACAGCGCCCTG TAGCGGGCGA TTAAGCGCGG CGGGTGTGGT 5520  
 GGTTACGGCG AGCGTGACCG CTACACTTGC CAGCGCCCTA CGCGCCGCTC CTTCGCTTT 5580  
 CTTCCTTCC TTCTCGCCA CGTTCGCCGG CTTCGCCCCGTAAGCTCTAA ATCGGGGGCT 5640  
 CCCCTTAGGG TTCCGATTAA GTGCTTACG GCACCTCGAC CCCAAAAAAC TTGATTTCGG 5700  
 TGATGGTCA CGTAGTGGGC CATCGCCCTG ATAGACGGTT TTTCGCCCC TTGACGTTGGA 5760  
 GTCCACGTTC TTTAATAGTG GACTCTTGTG CCAAACGTGA ACAACACTCA ACCCTATCTC 5820  
 GGGCTATTCT TTGATTAT AAGGGATTTT GCCGATTTCG GAACCACCAT CAAACAGGAT 5880  
 TTTCGCGTGC TGGGGCAAAC CAGCGTGGAC CGCTTGTGTC AACTCTCTCA GGGCCAGGCG 5940  
 GTGAAGGGCA ATCAGCTGTT GCGCGTCTCG CTGGTGAAAA GAAAAACAC CCTGGGGCCC 6000  
 AATACGCCAA CGCGCTCTCC CGCGCGTTG CGCGATTTCAT TAATGCAGCT GGCACGACAG 6060  
 GTTTCGGAC TCGAAAGCGG CGAGTGAGGG CAACGCCAATT AATGTGAGTT AGCTCACTCA 6120  
 TTAGGCACCC CAGGCTTAC ACTTTATGCT TCCGGCTCGT ATGTTGTGTC GAATTGTGAG 6180  
 CGGATAACAA TTTCACACGC CAAGGAGACA GTCATAATGA AATACCTATT GCCTACGGCA 6240  
 CGCGCTGGAT TGTATTACT CGCTGCCAA CCAGCCATGG CCGAGCTCTT CCCGCCATCT 6300  
 GATGAGCAGT TGAATCTGG AACTGCCCT GTTGTGTGCC TGCTGAATAA TTCTATCCC 6360  
 AGAGAGGCCA AAGTACAGTC GAAGGTGGAT AACGCCCTCC AATCGGGTAA CTCCCAGGAG 6420  
 AGTGTACAG ACCAGGACAG CAAGGACAGC ACCTACAGCC TCAGGAGCAC CCTGACCGTG 6480  
 AGCAAAGCAG ACTACGGAGAA ACACAAAGTC TACGCCCTGG AAGTCACCCCA TCAGGGCTG 6540  
 AGCTCGCCCG TCACAAAGAG CTTCAACAGG GGAGAGTGTGTT CTAGAACCGG TCACTTGGCA 6600  
 CTGGCCGTGG TTTTACAACG TCGTGACTGG GAAAACCCCTG CGCTTACCCCA AGCTTAATCG 6660  
 CCTTGCAGAA TTCCCTTTCG CGAGCTGGCG TAATAGCGAA GAGGCCCCGA CGCATCGCCC 6720  
 TTCCCAACAG TTGCGGAGCC TGAATGGCGA ATGGCGCTTT GCCTGGTTTC CGGCACCAAGA 6780  
 AGCGGTGCCG GAAAGCTGGC TGGAGTGGCA TCTTCCTGAG GCCGATACGG TCGTGGTCCC 6840  
 CTCAAACAG CAGATGCACG GTTACGATGC CCCATCTAC ACCAACGTAA CCTATCCCAT 6900  
 TACGGTCAAT CGGCCGTTCG TTCCCAACCGA GAATCCGAGG GGTGTTACT CGCTCACATT 6960  
 TAATGTGAT GAAAGCTGGC TACAGGAAGG CGAGACCGGA ATTATTTTG ATGGCGTTCC 7020  
 TATTGGTTAA AAAATGAGCT GATTAACAA AAATTTAACG CGAATTTAA CAAAATATTA 7080  
 ACGTTTACAA TTAAATATT TGCTTATACA ATCTTCCTGT TTTTGGGGCT TTCTGATTA 7140  
 TCAACCGGGG TACATATGAT TGACATGCTA GTTTACGAT TACCGTTCAT CGATTCTCTT 7200  
 GTTTCGCTCCA GACTCTCAGG CAATGACCTG ATAGCCTTTCG TAGATCTCTC AAAAATAGCT 7260  
 ACCCTCTCCG GCATTAATTATCAGCTAGA ACCGTTGAAT ATCATATTGA TGGTGATTTG 7320  
 ACTGTCTCCG GCCTTCTCA CCCTTGTAA TCTTACCTA CACATTACTC AGGCATTGCA 7380

TTTAAAATAT ATCAGGGTTC TAAAAATTCT TATCCTTGGG TTCAAATAAA GGCTTCTCCC 7440  
 GCAAAAGTAT TACAGGGTCA TAATGTTTT GGTACAACCC ATTAGCTTT ATGCTCTGAG 7500  
 CCTTTATTGC TTAATTTCGC TAATTCTTGT CCTTGCGCTGT ATGATTIATT CGATGTT 7557

## (2) INFORMATION FOR SEQ ID NO:5:

- (1) SEQUENCE CHARACTERISTICS:  
 (A) LENGTH: 8118 base pairs  
 (B) TYPE: nucleic acid  
 (C) STRANDEDNESS: both  
 (D) TOPOLOGY: circular

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO:5:

AATGCTACTA CTATTAGTAG AATTGATGCC ACCTTTCAAG CTCGGCCCC AAATGAAAAT	60
ATAGCTAAAC AGGTTATTGA CCATTGCGA AATGTATCTA ATGGTCAAAC TAAATCTACT	120
CGTTCCGAGA ATTGGGAATC AACTGTTACA TGGAATGAAA CTTCCAGACA CCGTACTTTA	180
GTGGCATATT TAAAACATGT TGAGCTACAG CACCAGATTG AGCAATTAAG CTCTAAGCCA	240
TCTGCAAAAA TGACCTCTTA TCAAAAGGAG CAATTAAAGG TACTCTCTAA TCCTGACCTG	300
TTGGAGTTTG CTICCGGTCT GGTCGCTT GAAGCTCGAA TTAAAACGCG ATATTGAAAG	360
TCTTTCGGGC TTCCCTTTAA TCTTTTGAT GCAATCCGCT TTGCTCTGA CTATAATAGT	420
CAGGGTAAAG ACCTGATTT TGATTTATGG TCATTCTCGT TTTCTGAACG GTTAAAGCA	480
TTTGAGGGGG ATTCAATGAA TATTTATGAC GATTCCGCAG TATTGGACGC TATCCAGTCT	540
AAACATTTTA CTATTACCCC CTCTGGAAA ACTTCTTTG CAAAAGCCTC TCGCTATTTC	600
CGTTTTATC GTCGTCTGGT AAACGAGGGT TATGATAGTG TTGCTCTTAC TATCCCTCGT	660
AATTCCCTTT GGCGTTATGT ATCTGGATTA GTTGAATGTG GTATTCTAA ATCTCAACTG	720
ATGAATCTTT CTACCTGTAA TAATGTTGTT CCGTTAGTTG GTTTTATTAA CGTAGATTTT	780
TCTTCCCAAC GTCCGTACTG GTATAATGAG CCAGTTCTTA AAATCCGATA AGGTAATTCA	840
CAATGATTAA AGTTGAAATT AAACCATCTC AAGCCCAATT TACTACTCGT TCTGGTGT	900
CTCGTCAGGG CAAGCCTTAT TCACTGAATG AGCAGCTTG TTACGGTGTAT TTGGGTAATG	960
AATATCCGGT TCTTGTCAAG ATTACTCTTG ATGAAGGTCA GCCAGCCTAT GCGCCTGGTC	1020
TGTACACCGT TCATCTGTCC TCTTTCAAAG TTGGTCAGTT CGGTTCCCTT ATGATTGACC	1080
GTCTGGCCT CGTTCCGGCT AAGTAACATG GAGCAGGTGCG CGGATTTCGA CACAATTAT	1140
CAGGGCGATGA TACAAATCTC CGTTGTACTT TGTTCGGCC TTGGTATAAT CGCTGGGGT	1200
CAAAGATGAG TGTGTTAGTG TATTCTTCCG CCTCTTTCGT TTTAGGTTGG TGCCTTCGTA	1260
GTGGCATTAC GTATTTTACCG CGTTTAATGG AAACCTCCTC ATGAAAAAGT CTTAGTCCT	1320
CAAAGCCTCT GTAGCCGTG CTACCCCTCGT TCCGATGCTG TCTTTGGCTG CTGAGGGTGA	1380
CGATCCCGCA AAAGCGGGCT TTAACCTCCCT GCAAGCCTCA GCGACCGAAT ATATCCGTTA	1440
TGGCTGGGGCG ATGGTTGTTG TCATTGTCCG CGCAACTATC GGTATCAAGC TGTGTTAAGAA	1500

ATTCACCTCG	AAAGCAAGCT	GATAAACCGA	TACAATTAAA	GGCTCCCTTT	GGAGCCTTTT	1560
TTTTGGAGA	TTTCAACGT	AAAAAAATTA	TTATTCGAA	TTCCCTTAGT	TGTTCCCTTC	1620
TATTCTCACT	CCGCTGAAAC	TGTTGAAAGT	TGTTTAGCAA	AACCCCATAAC	AGAAAATTCA	1680
TTTACTAACG	TCTGGAAAGA	CGACAAAAGT	TTAGATCGTT	ACGCTAACTA	TGAGGGTTGT	1740
CTGTGGAATG	CTACAGGCCT	TGTAGTTTGT	ACTGGTGACG	AAACTCAGTG	TTACGGTACA	1800
TGGGTTCTA	TTGGGCTTGC	TATCCCTGAA	AATGAGGGTG	GTGGCTCTGA	GGGTGGGGT	1860
TCTGAGGGTG	CGGGTTCTGA	GGGTGGGGT	ACTAAACCTC	CTGAGTACGG	TGATACACCT	1920
ATTCCGGGCT	ATACTTATAT	CAACCCCTCTC	GACGGCACTT	ATCCGGCTGG	TACTGACCAA	1980
AACCCCGCTA	ATCCTAATCC	TTCTCTTGAG	GAGTCTCAGG	CTCTTAATAC	TTTCATGTTT	2040
CAGAATAATA	GGTCCGAAA	TAGGCAGGGG	GCATTAACTG	TTTATACGGG	CACTGTTACT	2100
CAAGGCACTG	ACCCCGTTAA	AACTTATTAC	CACTACACTC	CTGTATCATC	AAAAGCCATG	2160
TATGACGCTT	ACTGGAACGG	TAAATTCAAGA	GAUTGCGCTT	TCCATTCTGG	CTTTAATGAA	2220
GATCCATTG	TTTGTGAATA	TCAAGGCCAA	TCGTCTGACC	TGCCTCAACC	TCCTGTCAAT	2280
GCTGGCGGCG	GCTCTGGTGG	TGGTTCTGGT	GGCGGCTCTG	AGGGTGGTGG	CTCTGAGGGT	2340
GGCGGTTCTG	AGGGTGGCGG	CTCTGAGGGA	GGCGGTTCCG	GTGGTGGCTC	TGGTTCCGGT	2400
GATTTTGATT	ATGAAAAGAT	GGCAAACGCT	AATAAGGGGG	CTATGACCGA	AAATGCCGAT	2460
AAAAACCCGC	TACAGTCTGA	CGCTAAAGGC	AAACTTGATT	CTGTCGCTAC	TGATTACGGT	2520
GCTGCTATCG	ATGGTTTCAT	TGGTCACGTT	TCCGGCCTTG	CTAATGGTAA	TGGTGCTACT	2580
GGTGATTITG	CTGGCTCTAA	TTCCCAAATG	GCTCAAGTCC	GTGACCGTGA	TAATTCACCT	2640
TTAATGAATA	ATTTCCGTCA	ATATTTACCT	TCCCTCCCTC	AATCGGTGAA	ATGTCGCCCT	2700
TTTGTCTTTA	GGCGTGGTAA	ACCATATGAA	TTTCTATTG	ATTGTACCAA	AATAAAACTTA	2760
TTCCGTGGTG	TCTTIGCGTT	TCTTTTATAT	GTTGCCACCT	TTATGTATGT	ATTTTCTACG	2820
TTTGCTAACAA	TACTGGCTAA	TAAGGAGTCT	TAATCATGCC	AGTTCCTTTG	GGTATTCCGT	2880
TATTATTGCG	TTTCTCTGGT	TTCCCTCTGG	TAACTTTGTG	CGGCTATCTG	CTTACTTTTC	2940
TTAAAAAGGG	CTTCGGTAAG	ATAGCTATTG	CTATTCATT	GTTCCTTGCT	CTTATTATTG	3000
GGCTTAACTC	AATTCTTGTG	GGTTATCTCT	CTGATATTAG	CGCTCAATTAA	CCCTCTGACT	3060
TTGTTCAAGGG	TGTTCAAGTTA	ATTCTCCCGT	CTAATGCCGT	TCCCTGTGTTT	TATGTTATTG	3120
TCTCTGTAAA	GGCTGCTATT	TTCACTTTTG	ACGTTAAACAA	AAAAATCGTT	TCTTATTG	3180
ATTGGGATAAA	ATAATATGGC	TGTTTATTTT	GTAACTGGCA	AATTAGGCTC	TGAAAGACGG	3240
CTCGTTAGCG	TTGGTAAGAT	TCAGGATAAA	ATTGTAGCTG	GGTCCAAAAT	AGCAACTAAT	3300
CTTGATTITAA	CGCTTCAAAA	CCTCCCCCAA	GTGGGGAGGT	TCGCTAAAAC	GCCTCGCGTT	3360
CTTACAATAC	CGGATAAGCC	TTCTATATCT	GATTGCTTG	CTATTGGCCG	CGGTAATGAT	3420
TCCTACGATG	AAAATAAAAA	CGGCTTGCTT	GTTCCTCGATG	AGTGGGGTAC	TTGGTTAAT	3480
ACCCGTTCTT	CGAATGATAA	GGAAAGACAG	CCGATTATTG	ATTGGTTTCT	ACATGCTCGT	3540

AAATTAGGAT CGGATATTAT TTTTCTTGT CAGGACTTAT CTATTGTTGA TAAACAGGCC 3600  
 CGTTCTGCAT TAGCTGAACA TGTTGTTAT TGTCTCGTC TGGACAGAAT TACTTACCT 3660  
 TTGTCGGTA CTTTATATTTC TCTTATTACT GGCTCGAAAA TGCCTCTGCC TAAATTACAT 3720  
 GTTGGCGTTG TTAAATATGG CGATTCTCAA TTAAGCCCTA CTGTTGAGCG TTGGCTTAT 3780  
 ACTGGTAAGA ATTTGTATAA CGCATATGAT ACTAACACAGG CTTTTCTAG TAATTATGAT 3840  
 TCCGGTGTATTT ATTCTTATTT AACGCCCTAT TTATCAGACGG GTCGGTATTT CAAACCATT 3900  
 AATTTAGGTC AGAACATGAA GCTTACTAAA ATATATTGAA AAAAGTTTTC ACGCGTTCTT 3960  
 TGTCTTGCCTA TTGGATTGTC ATCAGCATT ACATATAGT ATATAACCCA ACCTAAGCCG 4020  
 GAGGTAAAAA AGGTAGTCTC TCAGACCTAT GATTTGATA AATTCACTAT TGACTCTTCT 4080  
 CAGCGTCTTA ATCTAAGCTA TCGCTATGTT TTCAAGGATT CTAAGGAAA ATTAATTAAT 4140  
 AGCGACGATT TACAGAAGCA AGGTTATTCA CTCACATATA TTGATTTATG TACTGTTCC 4200  
 ATTAAAAAAAG GTAATTCAAA TCAAATTGTT AAATGTAATT AATTGTTTT TCTTGATGTT 4260  
 TGTTTCATCA TCTTCTTGT CTCAGGTAAT TGAAATGAAT AATTGGCTC TGCGCGATT 4320  
 TGTAACTTGG TATTCAAAGC AATCAGGCCA ATCCGTTATT GTTTCTCCCG ATGAAAAGG 4380  
 TACTGTTACT GTATATTCACT CTGACGTTAA ACCTGAAAAT CTACCGAATT TCTTATTT 4440  
 TGTTTACGT GCTAATAATT TTGATATGGT TGGTTCAATT CCTTCCATAA TTCAGAAGTA 4500  
 TAATCCAAAC AATCAGGATT ATATTGATGA ATTGCCATCA TCTGATAATC AGGAATATCA 4560  
 TGATAATTCC GCTCCTTCTG GTGGTTCTT TCTTCCGCAA AATGATAATG TTACTCAAAC 4620  
 TTTTAAATT AATAACGTTG GGGCAAAGGA TTTAATACGA CTGTCGAAT TGTTGTTAAA 4680  
 GTCTAATACT TCTAAATCCT CAAATGTATT ATCTATTGAC GGCTCTAATC TATTAGTTGT 4740  
 TAGTGCACCT AAAGATATT TAGATAACCT TCTCAATTCT CTTCTACTG TTGATTTGCC 4800  
 AACTGACCAAG ATATTGATTG AGGGTTGAT ATITGAGGTT CAGCAAGGTG ATGCTTACA 4860  
 TTTTCATTG GCTGCTGGCT CTCAGCGTGG CACTGTTGCA GGGGGTGTAA ATACTGACCG 4920  
 CCTCACCTCT GTTTTATCTT CTGCTGGTGG TTCTGTCGGT ATTTTAATG GCGATGTTT 4980  
 AGGGCTATCA GTTGGCGCAT TAAAGACTAA TAGCCATTCA AAAATATTGT CTGTGCCACG 5040  
 TATTCTTACG CTTCAAGGTC AGAAGGGTTC TATCTCTGTT GGCGACAAATG TCCCTTTAT 5100  
 TACTGGTCGT GTGACTGGTG AATCTCCAA TGAAATAAT CCATTCAGA CGATTGAGCG 5160  
 TCAAAATGTA GGTATTTCCA TGAGCGTTTT TCCTGTTGCA ATGGCTGGCG GTAATATTGT 5220  
 TCTGGATATT ACCACCAAGG CCGATAGTTT GAGTCTTCT ACTCAGGCAA GTGATGTTAT 5280  
 TACTAATCAA AGAAGTATTG CTACAACGGT TAATTGCGT GATGGACAGA CTCTTTACT 5340  
 CGGTGGCCTC ACTGATTATA AAAACACTTC TCAAGATTCT GGCGTACCGT TCCCTGTCAA 5400  
 AATCCCTTA ATCGGCCCTCC TGTGTTAGGTC CCGCTCTGAT TCCAACGAGG AAAGCACGTT 5460  
 ATACGTGCTC GTCAAAGCAA CCATAGTAGG CGCCCTGTAG CGGGCCATTAA AGCGCGGGCGG 5520  
 GTGTGGTGGT TACGGCGAGC GTGACCGCTA CACTTGGCAG CGCCCTAGCG CCCGCTCCTT 5580

TCGGCTTCTT	CCCTTCCTTT	CTCGCCACGT	TCGCCGGCTT	TCCCCGTCAA	GCTCTAAATC	5640
GGGGGCTCCC	TTAGGGTTC	CGATTAGTG	CTTACGGCA	CCTCGACCCC	AAAAAACTTG	5700
ATTTGGTGA	TCGTTCACGT	AGTGGGCCAT	CGCCCTGATA	GACGGTTTTT	CGCCCTTTGA	5760
CGTTGGAGTC	CACGGTCTTT	AATAGTGGAC	TCTTGTCCA	AACTGGAACA	ACACTCAACC	5820
CTATCTCGGG	CTATTCTTT	GATTATAAG	GGATTTGCC	GATTTCGGAA	CCACCATCAA	5880
ACAGGATTTT	CGCCTGCTGG	GGCAAACCGAG	CGTGGACCCGC	TTGCTGCAAC	TCTCTCAGGG	5940
CCAGGGCGTC	AAGGGCAATC	AGCTGTTGCC	CGTCTCGCTG	GTGAAAAGAA	AAACCACCCCT	6000
GGCGCCCAAT	ACGCAAACCG	CCTCTCCCCG	CGCGTTGGCC	GATTCAATTAA	TGGAGCTGGC	6060
ACGACAGGTT	TCCCGACTGG	AAAGCGGGCA	GTGAGCGCAA	CGCAATTAAAT	GTGAGTTAGC	6120
TCACTCAATTAA	GGCACCCAG	GCTTTACACT	TTATGCTTCC	GGCTCGTATG	TTGTGTGAA	6180
TTGTGAGCGG	ATAACAATTTC	CACACGCCAA	GGAGACAGTC	ATAATGAAAT	ACCTATTGCC	6240
TACGGCAGCC	GCTGGATTGT	TATTACTCGC	TGCCCAACCA	GCCATGGCCG	AGCTCTTCCC	6300
GCCATCTGAT	GAGCAGTTGA	AATCTGGAAC	TGCCCTGTGTT	GTGTGCTGC	TGAATAACTT	6360
CTATCCCAGA	GAGGCCAAAG	TACAGTGGAA	GGTGGATAAC	GCCCTCCAAT	CGGGTAACCTC	6420
CCAGGAGAGT	GTCACAGAGC	AGGACAGCAA	GGACACCCACC	TACAGCCTCA	GGAGCACCCCT	6480
GACGCTGAGC	AAAGCAGACT	ACGAGAAACA	CAAAGTCTAC	GCCTGCGAAG	TCACCCATCA	6540
GGGCCTGAGC	TCGGCCGTCA	CAAAGAGCTT	CAACAGGGGA	GAGTGTCTA	GAACGCGTCA	6600
CTTGGCACTG	GGCGTCGTTT	TACAACGTGCG	TGACTGGAA	AACCTGGCC	TTACCCAAGC	6660
TTTGTACATG	GAGAAAATAA	AGTCAAACAA	AGCACTATTG	CACTGGCACT	CTTACCGTTA	6720
CTGTTAACCC	CTGTGGAAA	AGCCGCCTCC	ACCAAGGGCC	CATCGGTCTT	CCCCCTGGCA	6780
CCCTCCTCCA	AGAGCACCTC	TGGGGCACA	GCGGCCCTGG	GCTCCCTGGT	CAAGACTAAT	6840
TCCCCGAACC	GGTGACGGTG	TCGTGGAACT	CAGGCCCTCT	GACCAGCCGC	GTGCACACCT	6900
TCCCGGCTGT	CCTACAGTCC	TCAGGACTCT	ACTCCCTCAG	CAGCGTGGTG	ACCGTGCCCT	6960
CCAGCAGCTT	GGGCACCCAG	ACCTACATCT	GCAACGTGAA	TCACAAGCCC	AGCAACACCA	7020
AGGTGGACAA	GAAAGCAGAG	CCCAAATCTT	GTACTAGTGG	ATCCTAACCG	TAACGACGTC	7080
CGGACTACGC	TTCTTAGGCT	GAAGGCGATG	ACCTGCTAA	GGCTGCAITC	AATAGTTTAC	7140
AGGCAAGTGC	TACTGAGTAC	ATGGCTACG	CTGGGCTAT	GGTAGTAGTT	ATAGTTGGTG	7200
CTACCATAGG	GATTAAATTAA	TTCAAAAAGT	TTACCGAGCAA	GGCTTCTTAA	GCAATAGCGA	7260
AGAGGCCCGC	ACCGATCGCC	CTTCCCAACA	GTTCGGCAGC	CTGAATGGCG	AATGGCGCTT	7320
TGCCCTGGTTT	CGGGCACCAG	AAGCGGTGCC	GGAAAGCTGG	CTGGAGTGGC	ATCTTCTGA	7380
GGCCGATACG	GTCCGTGTCC	CCTCAAACG	GCAGATGCAC	GGTTACGATG	CGCCCATCTA	7440
CACCAACGTA	ACCTATCCCA	TTACGGTCAA	TCCGGCTTT	GTTCGGACGG	AGAATCCGAC	7500
GGGTGTTAC	TCGCTCACAT	TTAATGTTGA	TGAAAGCTGG	CTACAGGAAG	GCCAGACCCG	7560
AATTATTTT	GATGGCGTTC	CTATTGGTTA	AAAAATGAGC	TGATTTAACCA	AAAATTAAAC	7620

GGGAATTITA ACAAAATATT AACGTTTACA ATTAAAATAT TTGCTTATAC AATCTTCCTG	7680
TTTTGGGGC TTTCTGATT ATCAACCGGG GTACATATGA TTGACATGCT AGTTTACGA	7740
TTACCGTTCA TCGATTCTCT TGTTTGTCC AGACTCTCAG GCAATGACCT GATAGCCTTT	7800
GTAGATCTCT CAAAAATAGC TACCCTCTCC GGCATTAATT TATCAGCTAG AACGGTGAA	7860
TATCATATTG ATGGTGATTG GACTGTCTCC GCCCTTTCTC ACCCTTTGGA ATCTTACCT	7920
ACACATTACT CAGGCATTGC ATTAAAATA TATGAGGGTT CTAAAATTT TTATCCTTGC	7980
GTTCAAATAA AGGCTTCTCC CGCAAAAGTA TTACAGGGTC ATAATGTTT TGGTACAACC	8040
GATTTAGCTT TATGCTCTGA GGCTTTATTG CTTAATTITG CTAATTCTTT GCCTTGCCTG	8100
TATGATTAT TGGACGTT	8118

## (2) INFORMATION FOR SEQ ID NO:6:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 22 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

## (ix) FEATURE:

- (A) NAME/KEY: misc\_difference
- (B) LOCATION: replace(5, "")
- (D) OTHER INFORMATION: /note- "S REPRESENTS EQUAL MIXTURE OF G AND C"

## (ix) FEATURE:

- (A) NAME/KEY: misc\_difference
- (B) LOCATION: replace(6, "")
- (D) OTHER INFORMATION: /note- "M REPRESENTS EQUAL MIXTURE OF A AND C"

## (ix) FEATURE:

- (A) NAME/KEY: misc\_difference
- (B) LOCATION: replace(8, "")
- (D) OTHER INFORMATION: /note- "R REPRESENTS EQUAL MIXTURE OF A AND G"

## (ix) FEATURE:

- (A) NAME/KEY: misc\_difference
- (B) LOCATION: replace(11, "")
- (D) OTHER INFORMATION: /note- "K REPRESENTS EQUAL MIXTURE OF G AND T"

## (ix) FEATURE:

- (A) NAME/KEY: misc\_difference
- (B) LOCATION: replace(20, "")
- (D) OTHER INFORMATION: /note- "W REPRESENTS EQUAL MIXTURE OF A AND T"

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO:6:

AGGTSMARCT KCTCGAGTCW GG

## (2) INFORMATION FOR SEQ ID NO:7:

- (i) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 22 base pairs
  - (B) TYPE: nucleic acid
  - (C) STRANDEDNESS: single
  - (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:7:

AGGTCCAGCT GCTCGAGTCT GG

## (2) INFORMATION FOR SEQ ID NO:8:

- (i) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 22 base pairs
  - (B) TYPE: nucleic acid
  - (C) STRANDEDNESS: single
  - (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:8:

AGGTCCAGCT GCTCGAGTCA GG

## (2) INFORMATION FOR SEQ ID NO:9:

- (i) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 22 base pairs
  - (B) TYPE: nucleic acid
  - (C) STRANDEDNESS: single
  - (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:9:

AGGTCCAGCT TCTCGAGTCT GG

## (2) INFORMATION FOR SEQ ID NO:10:

- (i) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 22 base pairs
  - (B) TYPE: nucleic acid
  - (C) STRANDEDNESS: single
  - (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:10:

AGGTCCAGCT TCTCGAGTCA GG

## (2) INFORMATION FOR SEQ ID NO:11:

- (i) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 22 bas pairs
  - (B) TYPE: nucleic acid
  - (C) STRANDEDNESS: single
  - (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:11:

AGGTCCAAT GCTCGAGTCT GG

(2) INFORMATION FOR SEQ ID NO:12:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 22 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:12:

AGGTCCAAT GCTCGAGTCA GG

(2) INFORMATION FOR SEQ ID NO:13:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 22 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:13:

AGGTCCAAT TCTCGAGTCT GG

(2) INFORMATION FOR SEQ ID NO:14:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 22 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:14:

AGGTCCAAGT TCTCGAGTCA GG

(2) INFORMATION FOR SEQ ID NO:15:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 22 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ix) FEATURE:

- (A) NAME/KEY: misc\_difference
- (B) LOCATION: replace(5..6, "")
- (D) OTHER INFORMATION: /note- "N-INOSINE"

(ix) FEATURE:

- (A) NAME/KEY: misc\_difference
- (B) LOCATION: replace(8, "")
- (D) OTHER INFORMATION: /note- "N-INOSINE"

## (ix) FEATURE:

- (A) NAME/KEY: misc\_difference
- (B) LOCATION: replace(11, "")
- (D) OTHER INFORMATION: /note- "N-INOSINE"

## (ix) FEATURE:

- (A) NAME/KEY: misc\_difference
- (B) LOCATION: replace(20, "")
- (D) OTHER INFORMATION: /note- "W REPRESENTS EQUAL MIXTURE OF A AND T"

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO:15:

AGGTNNANCT NCTCGAGTCW GG

22

## (2) INFORMATION FOR SEQ ID NO:16:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 38 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO:16:

CTATTAACTA GTAACGGTAA CAGTGGTGCC TTGCCCCA

38

## (2) INFORMATION FOR SEQ ID NO:17:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 30 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO:17:

AGGCTTACTA GTACAATCCC TGGGCACAAT

30

## (2) INFORMATION FOR SEQ ID NO:18:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 32 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO:18:

CCAGTTCCGA CCTCGTTGTG ACTCAGGAAT CT

32

## (2) INFORMATION FOR SEQ ID NO:19:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 32 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:19:

CCAGTTCCGA GCTCGTGTTC ACCCAGCCGC CC

(2) INFORMATION FOR SEQ ID NO:20:

32

(1) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 32 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:20:

CCAGTTCCGA GCTCGTGCTC ACCCAGTCTC CA

(2) INFORMATION FOR SEQ ID NO:21:

32

(1) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 32 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:21:

CCAGTTCCGA GCTCCAGATG ACCCAGTCTC CA

(2) INFORMATION FOR SEQ ID NO:22:

32

(1) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 32 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:22:

CCAGATGTGA GCTCGTGATG ACCCAGACTC CA

(2) INFORMATION FOR SEQ ID NO:23:

32

(1) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 32 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:23:

CCAGATGTGA GCTCGTCATG ACCCAGTCTC CA

(2) INFORMATION FOR SEQ ID NO:24:

32

(1) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 32 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:24:

CCAGTTCCGA GCTCGTGATG ACACAGTCTC CA

(2) INFORMATION FOR SEQ ID NO:25:

- (1) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 32 base pairs
  - (B) TYPE: nucleic acid
  - (C) STRANDEDNESS: single
  - (D) TOPOLOGY: linear

32

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:25:

GCAGCATTCT AGAGTTTCAG CTCCAGCTTG CC

(2) INFORMATION FOR SEQ ID NO:26:

- (1) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 34 base pairs
  - (B) TYPE: nucleic acid
  - (C) STRANDEDNESS: single
  - (D) TOPOLOGY: linear

32

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:26:

GCGCCGTCTA GAATTAACAC TCATTCTGT TGAA

(2) INFORMATION FOR SEQ ID NO:27:

- (1) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 37 base pairs
  - (B) TYPE: nucleic acid
  - (C) STRANDEDNESS: single
  - (D) TOPOLOGY: linear

34

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:27:

GATCCTAGGC TGAAGGCGAT GACCCTGCTA AGGCTGC

(2) INFORMATION FOR SEQ ID NO:28:

- (1) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 35 base pairs
  - (B) TYPE: nucleic acid
  - (C) STRANDEDNESS: single
  - (D) TOPOLOGY: linear

37

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:28:

ATTCAATAGT TTACAGGCAA GTGCTACTGA GTACA

35

## (2) INFORMATION FOR SEQ ID NO:29:

- (i) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 35 base pairs
  - (B) TYPE: nucleic acid
  - (C) STRANDEDNESS: single
  - (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:29:

TTGGCTACGC TTGGGCTATG GTAGTAGTTA TAGTT

35

## (2) INFORMATION FOR SEQ ID NO:30:

- (i) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 35 base pairs
  - (B) TYPE: nucleic acid
  - (C) STRANDEDNESS: single
  - (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:30:

GGTGCTACCA TAGGGATTAA ATTATTCAAA AAGTT

35

## (2) INFORMATION FOR SEQ ID NO:31:

- (i) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 18 base pairs
  - (B) TYPE: nucleic acid
  - (C) STRANDEDNESS: single
  - (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:31:

TACCGAGCAAG GCTTCTTA

18

## (2) INFORMATION FOR SEQ ID NO:32:

- (i) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 39 base pairs
  - (B) TYPE: nucleic acid
  - (C) STRANDEDNESS: single
  - (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:32:

AGCTTAAGAA GCCTTGCTCG TAAACTTTTT GAATAATT

39

## (2) INFORMATION FOR SEQ ID NO:33:

- (i) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 36 base pairs
  - (B) TYPE: nucleic acid
  - (C) STRANDEDNESS: single
  - (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:33:

AATCCCTATG GTAGCACCAA CTATAACTAC TACCAT

(2) INFORMATION FOR SEQ ID NO:34:

36

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 35 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:34:

AGCCCCAAGCG TAGCCAATGT ACTCAGTAGC ACTTG

(2) INFORMATION FOR SEQ ID NO:35:

35

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 34 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:35:

CCTGTAAACT ATTGAATGCA GCCTTAGCAG GGTC

(2) INFORMATION FOR SEQ ID NO:36:

34

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 16 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:36:

ATCCCCTTGA GCCTAG

(2) INFORMATION FOR SEQ ID NO:37:

16

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 21 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:37:

CATTTTTGCA GATGGCTTAG A

(2) INFORMATION FOR SEQ ID NO:38:

21

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 18 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:38:

TAGGATTAAC GTCCAATA

18

(2) INFORMATION FOR SEQ ID NO:39:

- (1) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 26 base pairs
  - (B) TYPE: nucleic acid
  - (C) STRANDEDNESS: single
  - (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:39:

ATATATTTTA GTAAGCTTCA TCTTCT

26

(2) INFORMATION FOR SEQ ID NO:40:

- (1) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 23 base pairs
  - (B) TYPE: nucleic acid
  - (C) STRANDEDNESS: single
  - (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:40:

GACAAAGAAC GCGTGAAAAC TTT

23

(2) INFORMATION FOR SEQ ID NO:41:

- (1) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 35 base pairs
  - (B) TYPE: nucleic acid
  - (C) STRANDEDNESS: single
  - (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:41:

GCGGGCCTCT TCGCTATTGC TTAAGAAGCC TTGCT

35

(2) INFORMATION FOR SEQ ID NO:42:

- (1) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 43 base pairs
  - (B) TYPE: nucleic acid
  - (C) STRANDEDNESS: single
  - (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:42:

AAACGACGGC CAGTGCCAAG TGACCCGTGT GAAATTGTTA TCC

43

## (2) INFORMATION FOR SEQ ID NO:43:

- (1) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 43 base pairs
  - (B) TYPE: nucleic acid
  - (C) STRANDEDNESS: single
  - (D) TOPOLOGY: linear

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO:43:

GGCGAAAGGG AATTCTGCAA GGCGATTAAG CTTGGGTAAC GCC

43

## (2) INFORMATION FOR SEQ ID NO:44:

- (1) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 36 base pairs
  - (B) TYPE: nucleic acid
  - (C) STRANDEDNESS: single
  - (D) TOPOLOGY: linear

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO:44:

GGCGTTACCC AAGCTTGTATGGAGAAA ATAAAG

36

## (2) INFORMATION FOR SEQ ID NO:45:

- (1) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 42 base pairs
  - (B) TYPE: nucleic acid
  - (C) STRANDEDNESS: single
  - (D) TOPOLOGY: linear

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO:45:

TGAAACAAAG CACTATTGCA CTGGCACTCT TACCGTTACC GT

42

## (2) INFORMATION FOR SEQ ID NO:46:

- (1) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 42 base pairs
  - (B) TYPE: nucleic acid
  - (C) STRANDEDNESS: single
  - (D) TOPOLOGY: linear

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO:46:

TACTGTTAAC CCCTGTGACA AAAGCCCCCC AGGTCCAGCT GC

42

## (2) INFORMATION FOR SEQ ID NO:47:

- (1) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 44 base pairs
  - (B) TYPE: nucleic acid
  - (C) STRANDEDNESS: single
  - (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:47:

TCCAGTCAGG CCTATTGTGC CCAGGGATTG TACTAGTGGAA TCCG

(2) INFORMATION FOR SEQ ID NO:48:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 38 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:48:

TGGCGAAAGG GAATTCGGAT CCACTAGTAC AATCCCTG

(2) INFORMATION FOR SEQ ID NO:49:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 42 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:49:

GGCACAAATAG GCCTGACTCG AGCAGCTGGA CCAGGGGGGC TT

(2) INFORMATION FOR SEQ ID NO:50:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 42 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:50:

TTGTCACAGG CGTAAACAGT AACGGTAACG GTAAGTGTGC CA

(2) INFORMATION FOR SEQ ID NO:51:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 42 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:51:

GTGCAATAGT GCTTTGTTC ACTTTATTTT CTCCATGTAC AA

(2) INFORMATION FOR SEQ ID NO:52:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 21 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:52:

TAACGGTAAG AGTGCCAGTG C

(2) INFORMATION FOR SEQ ID NO:53:

(1) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 32 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:53:

CACCTTCATG AATTGGCAA GGAGACAGTC AT

(2) INFORMATION FOR SEQ ID NO:54:

(1) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 22 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:54:

AATTGGCAA GGAGACAGTC AT

(2) INFORMATION FOR SEQ ID NO:55:

(1) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 39 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:55:

AATGAAATACTTGCCTA CGGCAGCCGC TGGATTGTT

(2) INFORMATION FOR SEQ ID NO:56:

(1) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 39 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:56:

ATTACTCGCT GCCCAACCAG CCATGGCCGA GCTCGTGAT

## (2) INFORMATION FOR SEQ ID NO:57:

- (i) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 39 base pairs
  - (B) TYPE: nucleic acid
  - (C) STRANDEDNESS: single
  - (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:57:

GACCGAGACT CCAGATATCC AACAGGAATG AGTGTAAAT

## (2) INFORMATION FOR SEQ ID NO:58:

39

- (i) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 13 base pairs
  - (B) TYPE: nucleic acid
  - (C) STRANDEDNESS: single
  - (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:58:

TCTAGAACGC GTC

## (2) INFORMATION FOR SEQ ID NO:59:

13

- (i) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 45 base pairs
  - (B) TYPE: nucleic acid
  - (C) STRANDEDNESS: single
  - (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:59:

TTCAGGTGAGC TTCTAGAATT AACACTCATT CCTGT

## (2) INFORMATION FOR SEQ ID NO:60:

45

- (i) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 39 base pairs
  - (B) TYPE: nucleic acid
  - (C) STRANDEDNESS: single
  - (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:60:

TGGATATCTG GAGTCTGGGT CATCACCGAGC TCGGCCATG

## (2) INFORMATION FOR SEQ ID NO:61:

39

- (i) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 39 base pairs
  - (B) TYPE: nucleic acid
  - (C) STRANDEDNESS: single
  - (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:61:

GCTGGTTGGG CAGCGAGTAA TAACAATCCA CGGGCTGCC

(2) INFORMATION FOR SEQ ID NO:62:

39

(1) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 37 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:62:

GTAGGCAATA GGTATTTCAT TATGACTGTC CTGGCC

(2) INFORMATION FOR SEQ ID NO:63:

37

(1) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 30 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:63:

TGACTGTCTC CTTGGCGTGT GAAATTGTTA

(2) INFORMATION FOR SEQ ID NO:64:

30

(1) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 36 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:64:

TAACACTCAT' TCCGGATGGA ATTCTGGACT CTGGGT

(2) INFORMATION FOR SEQ ID NO:65:

36

(1) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 24 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:65:

GCCAGTGCCA AGTGACGGCT TCTA

(2) INFORMATION FOR SEQ ID NO:66:

24

(1) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 26 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:66:

ATATATTITA GAAAGCTTCA TCTTCT

(2) INFORMATION FOR SEQ ID NO:67:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 23 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

26

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:67:

GACAAAGAAC GCGTGAAAAC TTT

(2) INFORMATION FOR SEQ ID NO:68:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 76 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

23

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:68:

CTGAACCTGT CTGGGACCAC AGTTGATGCT ATAGGATCAG ATCTAGAATT CATTAGAGA  
CTGGCCTGGC TTCTGC

60

(2) INFORMATION FOR SEQ ID NO:69:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 80 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

76

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:69:

TGGACCGTTG GTAGGAATAA TGCAATTAA GGAGTAGCTC TAAATTAGA ATTCACTAC  
ACCCAGTGCA TCCAGTAGCT

60

(2) INFORMATION FOR SEQ ID NO:70:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 27 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

80

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:70:

GGTAAACAGT AACGGTAAGA GTGCCAG

27

## (2) INFORMATION FOR SEQ ID NO:71:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 54 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO:71:

CGCCTTCAGC CTAAGAACGG TAGTCCGGAA CGTCGTACGG GTAGGATCCA CTAG

54

## (2) INFORMATION FOR SEQ ID NO:72:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 41 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO:72:

CACCGGTTCG GGGAAATTAGT CTTGACCAGG CAGCCCAGGG C

41

## (2) INFORMATION FOR SEQ ID NO:73:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 51 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO:73:

ATTCCACACA TTATACGAGC CGGAAGCATA AAGTGTCAAG CCTGGGGTGC C

51

## (2) INFORMATION FOR SEQ ID NO:74:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 42 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO:74:

CTGCTCATCA GATGGCGGGA AGAGCTCGGC CATGGCTGGT TG

42

## (2) INFORMATION FOR SEQ ID NO:75:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 42 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

78

(x1) SEQUENCE DESCRIPTION: SEQ ID NO:75:

GAACAGAGTG ACCGAGGGGG CGAGCTCCGC CATGGCTGGT TG

42